Functional Genomics of Streptomics

Borodina, Irina

Publication date:
2007

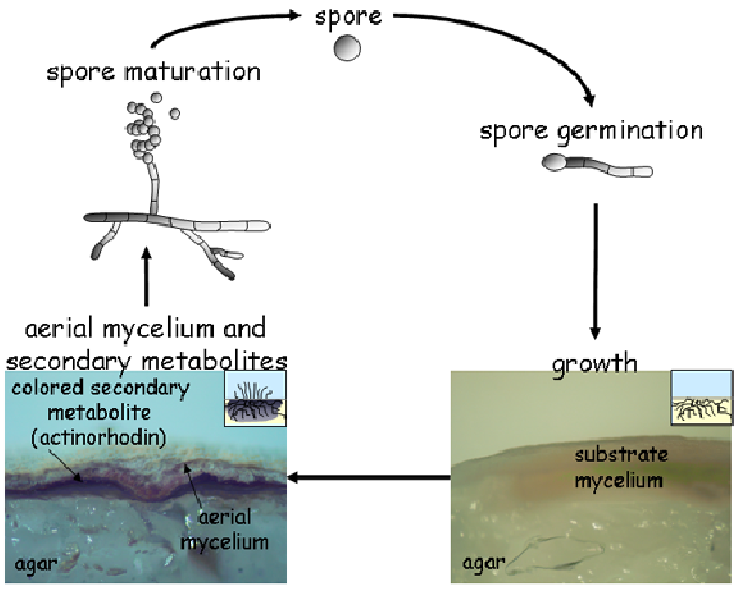
Document Version
Publisher's PDF, also known as Version of record

Functional Genomics of *Streptomyces*

Irina Borodina

Ph.D. Thesis

October 2007



BioCentrum-DTU
TECHNICAL UNIVERSITY OF DENMARK

Preface

This PhD study was carried out at Center for Microbial Biotechnology, BioCentrum-DTU in Technical University of Denmark from February 2004 till April 2007 and was sponsored by a DTU scholarship. I'm enormously grateful to the Technical University of Denmark and the Danish governmental scholarship programs, which allowed me to follow my dream of studying biotechnology first as an exchange student, then as an international master student and finally as a PhD student.
My warmest thanks and deepest appreciation go to my supervisor Prof. Jens Nielsen for being encouraging, supportive, inventive and always optimistic supervisor from a capital "S".
I'm very thankful to Anna Eliasson Lantz for help, advice and many hours of discussion, trying to solve the weird "*Streptomyces* problems".
Jochen Forster (Fluxome, Denmark) helped a lot with getting me started on stoichiometric models, in spite that he was busy at that time with his start-up company. I appreciate his help a lot.
During this project I was lucky to get many collaborations and I'm enormously grateful to all the people for their trust and support. This includes Preben Krabben (University College of London, UK), who with his truly encyclopaedic knowledge of Actinomycetes metabolism helped to build a high-quality *Streptomyces* model. Jeroen Siebring from Groningen University (Holland) entrusted me with characterization of his mutants and was a great partner to write a paper together. Mervyn Bibb and Nick Bird from John Innes Centre (UK) supplied me with materials for molecular biology of *Streptomyces* and helped out with troubleshooting. Mervyn also gave an idea to work on stringent response in *Streptomyces coelicolor*. Colin Smith welcomed me in his laboratory at Surrey University (UK) and set his microarrays facility at my disposal. During my microarrays quest Vasillis Mersinias and Giselda Bucca helped me to get through the practical problems and Emma Laing and Giles Velarde assisted in microarray data uploading and normalization. I am very happy that Prashant Bapat has joined our center in 2006 and officially became the second person working on *Streptomyces* at CMB. It was a great pleasure to brainstorm on new project ideas, work together in the lab and to discuss results with Prashant.
My super students Pichahpuk, Jie, Roberto and Dean, who taught me more about project management skills than one can learn in twenty courses, deserve special thanks.
And where would I be without the excellent technical support from Tina Johansen, Elisabeth Krøger, Martin Nielsen, Jette Mortensen, Bettina Jørgensen, and Hanne Christensen?
The most important, I want to thank all the colleagues at CMB for great academic and social environment.

Finally I thank my husband Asger and son Richard for the moral and practical support, without which this PhD. study would not be possible.

Content

“Science is facts; just as houses are made of stones, so is science made of facts; but a pile of stones is not a house and a collection of facts is not necessarily science.”

Henri Poincare
French mathematician & physicist (1854 - 1912)

Summary

Streptomyces are Gram-positive bacteria with some specific features as filamentous growth and morphological differentiation. They are among the most numerous and wide-spread soil bacteria, playing an important role in degradation of biological polymers like proteins, cellulose, hemicelluloses, chitin and others. *Streptomyces* are famous for making biologically active compounds: they produce more than half of naturally derived antibiotics in current use as well as anticancer drugs, immunosuppressants, antihelmintic agents, etc. These natural products are synthesized in complex secondary metabolites pathways and there is a lot of interest to understand how these pathways are regulated and what influences the secondary metabolites productivity. The ultimate aim is to decrease the production costs and to produce novel biologically active compounds.

To approach the problem we decided to use the genomic information available for the most studied *Streptomyces* species – *S. coelicolor* A3(2) – to connect genes to enzymatic function in the form of a genome-scale metabolic model. We see these models as a focal point of systems biology as they allow collection of various data types in a form suitable for mathematical analysis.

We reconstructed a genome-scale metabolic network of *Streptomyces coelicolor* A3(2) based on annotated genes, physiological and biochemical information. The stoichiometric model included 700 unique reactions and 500 metabolites, describing the function of 711 genes. In a comparative analysis with the *S. avermitilis* genome we showed that metabolic genes are highly conserved between *Streptomyces* species and therefore the model is suitable for use with other Streptomycetes. We used the model to assess metabolic capabilities for growth and polyketides production, to identify essential reactions, to predict knock-out phenotypes, and to identify gaps in genome annotation.

The model was useful in interpreting the flux data from isotope labeling experiments and for analyzing gene expression data. We employed it in integrated analysis of an overproducing *S. coelicolor* A3(2) mutant with deletion of one phosphofructokinase isoenzyme. We also applied the model for studying amino acids metabolism and stringent response in *S. coelicolor* A3(2) and its relaxed mutant.

Finally we used the model to predict knock-outs, which would lead to improved production of polyketides, and validated the best predictions experimentally.

Sammenfatning

Streptomyceter er Gram-positive bakterier, som har nogle unikke egenskaber såsom filamentøs vækst og morfologisk differentiering. Denne type bakterier findes udbredt i jord, hvor de spiller en vigtig rolle i nedbrydning af biologiske polymerer såsom proteiner, cellulose, hemicellulose og chitin. *Streptomyces* er kendte for at producere en række forskellige bioaktive stoffer, og mere end halvdelen af de antibiotika der er I brug i dag produceres af disse bakterier. Streptomyceter anvendes også til produktion af andre bioaktive stoffer der anvendes til kræftmedicin og imunohæmmende stoffer. Alle disse produkter er såkaldte sekundære metaboliter, som produceres af bakterierne via komplekse pathways. For at udnytte Streptomyceter kommercielt er det vigtigt at forstå hvordan disse pathways er reguleret og hvilke faktorer der påvirker cellernes produktivitet. Hermed kan bakterierne forbedres ved gensplejsning og produktionsomkostninger kan nedbringes og endvidere vil det være muligt at fremstille helt nye bioaktive stoffer.

I dette projekt er der opstillet en detaljeret matematisk model der beskriver en meget stor del af de mange forskellige pathways der opererer i den mest studerede *Streptomyces* art - *Streptomyces coelicolor*. Den opstillede model er baseret på information om hvilke enzymer der er kodet for i cellens genom, men anden information er også blevet anvendt i modelopstillingen. Foruden anvendelse til analyse af cellens stofskifte er modellen også vigtig i system biologi fordi den tillader at integrere analyse af forskellige –omic data typer.

Den metaboliske model af *Streptomyces coelicolor* A3(2) blev rekonstrueret ud fra genomsekvens og oplysninger om organismens fysiologi og biokemi. Den støkiometriske model indeholder 700 unikke reaktioner og 500 metaboliter og den beskriver funktioner af 711 gener. Ved at sammenligne *S. coelicolor* og *S. avermitilis* genomerne kunne vi bevise at metaboliske gener er meget bevarede mellem forskellige *Streptomyces* stammer og derfor kunne vores model bruges som udgangspunkt for opstilling af modeller for andre Streptomyceter. I løbet af model opbygning kunne vi endvidere tilskrive funktion til flere gener. Modellen er brugt til at beregne cellernes muligheder for at vokse og producere polyketider og til at finde de gener som er absolut nødvendige for vækst under forskellige betingelser.

Modellen er endvidere anvendt til at analysere flux data fra isotop mærkning forsøg og for at analysere gene ekspressionsdata. Modellen blev brugt i integreret analyse af en overproducerende *S. coelicolor* mutant, hvor et gen der koder for en phosphofruktokinase er slettet. Modellen blev også anvendt til at forstå aminosyremetabolismen og stringent respons i *S. coelicolor* A3(2).

Endelig er modellen brugt til at forudsige hvilke gener der skal deleteres for at forbedre polyketidproduktionen i *S. coelicolor*. De bedste varianter identificeret med modellen blev testet i laboratoriet.

CHAPTER 1

Introduction

Functional genomics is a rather recent branch in biology. It aims at understanding the functions of the genes from the vast wealth of information generated by post-genomic tools: genome-sequencing, gene expression analysis, proteomics, metabolomics and others.

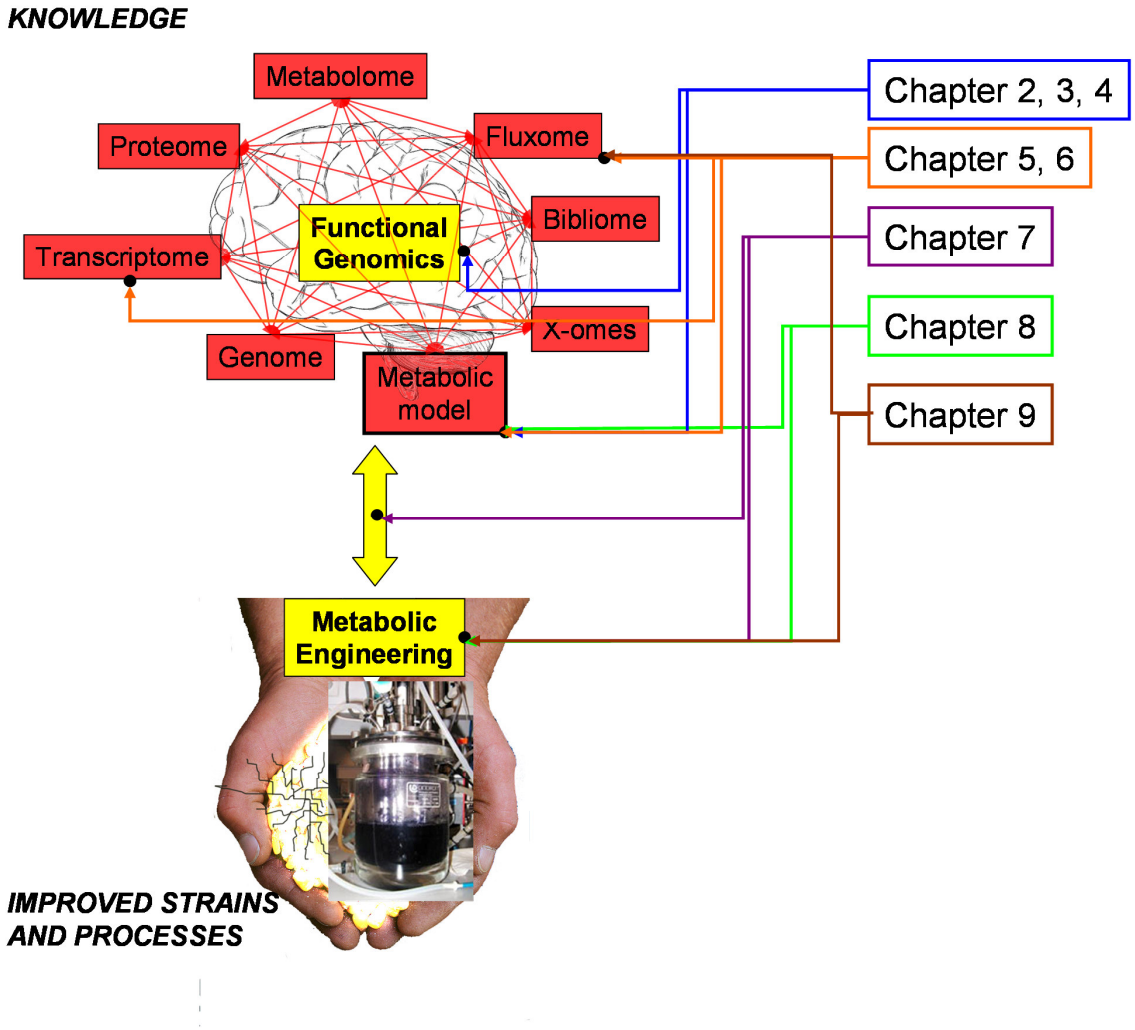
While perhaps the most important information we want to uncover is the role of the human genes, those are quite a challenge to study, both experimentally and from the ethical perspective. The studies of genes from the simplest organisms – bacteria – have provided us with immense amount of information both on genes function and on their regulation. This knowledge can be transferred to higher organisms thanks to evolutional conservation of DNA and protein sequences.

Streptomyces bacteria distinguish themselves by a high GC content, filamentous growth, complex regulation that allows morphological differentiation and sporulation and most of all by production on numerous bioactive secondary metabolites. Many different species of Streptomyces are used in industry for production of antibiotics, anticancer drugs, immunosupressants and others. One *Streptomyces* species – *S. coelicolor* (*lat.* heaven blue) has, however, been particularly in focus of academic research. The bacterium was chosen by Dr. David Hopwood, when he started his PhD. study in early 50's. He thought that the blue pigment would make a convenient marker so that mutants are easily detected from the changed color [1]. Until early 80s' the genetic research was mainly done by generating random mutants and mapping the mutations to certain locus of the chromosome. With the revolutionary discoveries of molecular biology techniques in early 80's the scientists could make more targeted changes in the genome. Still in the absence of genomic sequence, it could take months to map a mutation to a particular region in the chromosome and to identify the function of the gene. Some years later sequencing technology started to advance fast, making sequencing a more routine procedure. Finally, whole genomes could be sequenced, marking the entrance into a post genomic era. A real push to *Streptomyces* research was the sequencing of the *S. coelicolor* A3(2) genome in 2002 [2]. At that time this was the largest sequenced bacterial genome of more than 8 Mbp and with almost 8,000 open reading frames (ORFs), which were twice as many as in the *Escherichia coli* and also more than about 6,000 ORFs that have been identified in the eukaryote *Saccharomyces cerevisiae*. In was found that in *Streptomyces* a large number of metabolic genes were duplicated, but it's not known yet why. About 30% of the open reading frames did not have any match in the protein databases, so it is also a challenge to understand what these genes are doing.

Currently there is much interest in functional genomics of *S. coelicolor* as it is expected that this may lead to improving industrial processes but also improve our understanding of secondary metabolism in general and hereby help us to search for novel natural products (Figure 1-1). In this context a key issue is how the genetic information is translated to information concerning the physiology of the cells. This requires an integrated analysis of the cellular metabolism, and for this it is valuable to reconstruct the metabolic network and use the concept of metabolic flux balancing for analysis of how the network operates. A review in **Chapter 2** describes how metabolic networks help us to understand the cellular function, making an important part of functional genomics.

In this project the metabolic network of *S. coelicolor* was reconstructed using the annotated genome, pathway databases and knowledge about the organism's biochemistry and physiology (**Chapter 3**). Due to the large size of the genome of this organism and the many redundant metabolic genes identified it was very useful to map all the metabolic functions and use the reconstructed metabolic network in future functional studies. The reconstructed network was analyzed for its topology structure and other structural features. The *S. coelicolor* model was used as a template for quick reconstruction of a less studied parasitic *S. scabies*, recently sequenced, but not yet annotated (**Chapter 4**).

Figure 1-1. Connections between functional genomics and metabolic engineering and contribution of the thesis to different areas.



We used various post genomic techniques to characterize antibiotics overproduction caused by phosphofructokinase A2 deletion (**Chapter 5**) and to study the global role of stringent response mediator ppGpp (**Chapter 6**). Our model was helpful in data interpretation.

The knowledge accumulated in functional genomics studies can be used in metabolic engineering to rationally improve product formation or cellular properties of the producing organism. On the other hand, the results of the metabolic engineering add to our understanding of the cellular function. Therefore, functional genomics and metabolic engineering are tightly interconnected (Figure 1-1). We describe the strategies for metabolic engineering of Streptomycetes as well as application of the post-genomic data for choosing rational strategies in a review (**Chapter 7**). We demonstrate how metabolic model could be used for improving polyketides production in *S. coelicolor* (**Chapter 8**). We also show that detailed physiological characterization including isotope-based flux analysis is useful in guiding metabolic engineering strategies in industrial *Streptomyces* strains (**Chapter 9**).

References

1. Hopwood, D. A. 1999. "*Forty years of genetics with Streptomyces: from in vivo through in vitro to in silico*." Microbiology **145**:2183-2202.

2. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. 2002. "*Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).*" Nature **417**:141-147.

CHAPTER 2

From genomes to *in silico* cells via metabolic models

Irina Borodina and Jens Nielsen

Center for Microbial Biotechnology, BioCentrum-DTU, Building 223, DK-2800 Kgs. Lyngby, Denmark

Content of this chapter is published in: Current Opinion in Biotechnology, 2005, 16(3):350-355.

2.1 Summary

Genome-scale metabolic models are the focal point of systems biology as they allow collection of various data types in a form suitable for mathematical analysis. High-quality metabolic networks and metabolic networks with incorporated regulation have been successfully used for analysis of phenotypes from phenotypic arrays and gene deletion studies. They have also been applied for gene expression analysis guided by metabolic network structure, which lead to identification of genes commonly regulated due to connections on metabolites level. Thus, genome-scale metabolic modeling currently stands out as one of the most promising approaches to obtain an *in silico* prediction of cellular function based on interaction of all of the cellular components.

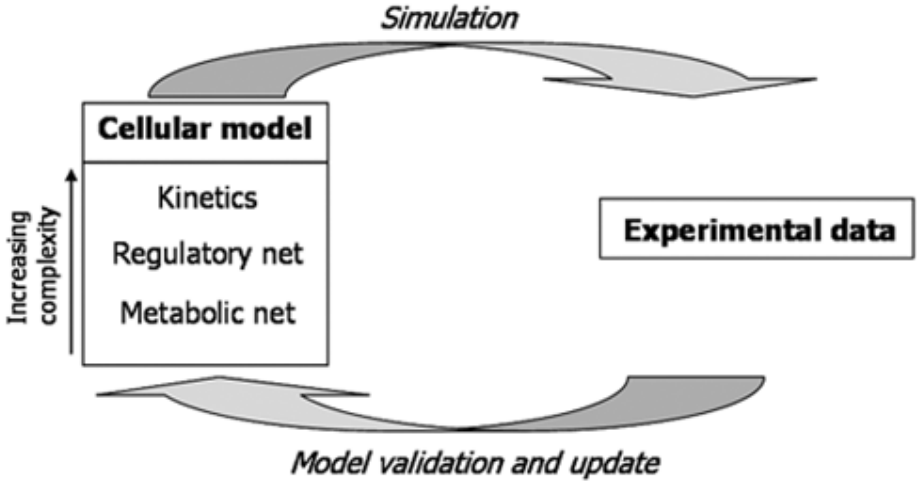
Keywords

Metabolic model, systems biology

2.2 Introduction

Systems biology is the integration of experimental and computational approaches in achieving its overall goal of explaining and predicting complex cellular behaviors of biological systems. It differs from traditional biology approaches in that it aims to study the interaction between the network components rather than the properties of the individual components themselves [1]. Using different 'omics' technologies a wealth of information is becoming available on the levels of many different cellular components and to some extent on the interaction between some of these components. Analysis and integration of these data will greatly benefit from the use of mathematical models, and cellular models are at the core of systems biology. *We therefore see systems biology as an iterative process where a model is developed and continuously improved through comparison with experimental data and integration of new features* (Figure 2-1).

Figure 2-1. Iterative process of cellular model development.



Mathematical models can vary in both complexity and approach, but in all cases the value of a mathematical model should be judged based on the objective of the

modeling exercise. Most common structures of models used for functional genomics and systems biology objectives are listed in Table 2-1.

Table 2-1. Objectives and types of models commonly used in systems biology

Objective	Type of model
Identification of significance	Statistical models used to perform e.g. t-test or Anova test
Grouping of information	Statistical models used for clustering of data
Mapping of interactions and development of predictive algorithms	Data driven models like neural network or Hidden Markov models
Quantitative analysis and dynamic simulation	Kinetic models that describe the dynamics of the individual components in the sub-system
Whole cell function described by function of a large fraction of the system components	Metabolic models with added regulatory features

Many different clustering methods and statistical tests have been developed to analyze 'omics' data, and these have been quite successful in structuring the information content. However, these mathematical methods (or models) are static of nature and they have no predictive power. Furthermore, even though they have proven valuable for data analysis they do not rely on the wealth of biochemical and molecular biology information gathered the last fifty years.
Neural networks and Hidden Markov models are more advanced approaches to data analysis, and using such models it is possible to develop algorithms with predictive power. Thus it is possible to identify interactions between the different components in the system, even during dynamic processes as illustrated recently in a study of the cell cycle in yeast [2,3]. Despite the value of these models for pathway reconstruction, these models are data driven and can therefore not be used to simulate e.g. the effect of different environmental conditions on cellular growth.
In order to simulate the dynamic response of biological systems to different stimuli or to different perturbations of the system, there are often applied kinetic models. These models are typically build from first principles, where the mass balances for the individual components are described by differential equations, and the kinetics of the different reactions and processes are described with more or less empirical rate equations. There are many examples of this type of models [4,5], but all models presented so far are for relatively small systems – at least in the perspective of a whole cell. Hence these models are basically not very useful for integration of omics data, and they are also not able to describe overall cellular function. In the long run it may, however, be possible to collect all the kinetics parameters and define rate equations for all reactions and processes in the cell, and hereby build a complete kinetic model for the cell.
Even though there are many more or less empirical models that can simulate growth, these models are not really relevant in the context of systems biology. There is, however, one type of models that have shown a surprising ability to simulate growth of living cells – namely metabolic models, or often referred to as stoichiometric flux balance models [6-8]. These models rely on balancing of all intracellular metabolites in the metabolic network, which basically represents the core machinery of cellular function as it provides the building blocks and Gibbs free energy for cellular synthesis. The balancing of metabolites supplies a large number of constraints, and through the use of linear programming it is possible to find solutions that obey a specified objective function. Growth optimization has proven to be a generally

applicable objective function, and it has been demonstrated that even if cells do not obey this objective function [9,10], e.g. for *Escherichia coli* cells growing on glycerol, they will adapt to this if they are evolved over many generations [11].

In this review we will discuss the process of metabolic model reconstruction and the recent advances in 'omic' data analysis guided by metabolic models.

2.3 Metabolic models: reconstruction

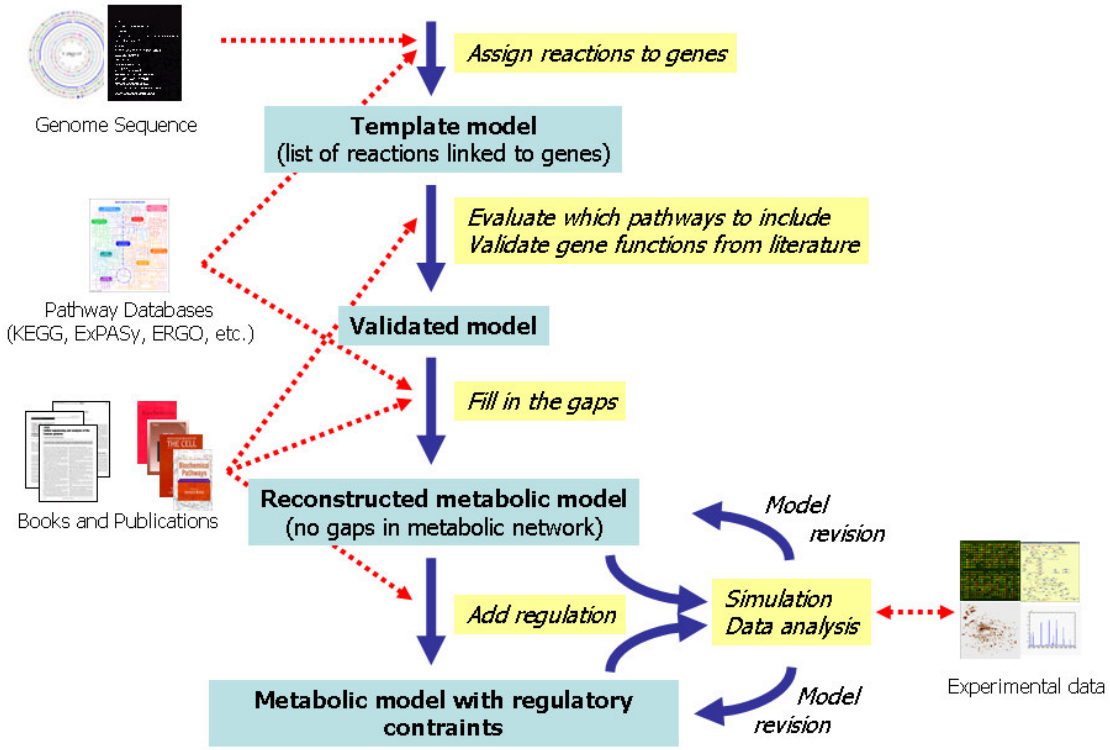
As mentioned above metabolic models currently represent the only attempt to construct mathematical models that include a large fraction of the components in the system and can simulate whole cellular behavior. There are a couple of reasons for why this may be the case:

- the metabolism is very well established for many organisms, and the model therefore rely on much information of biochemistry and at the same time it integrates information from many different sources.
- the concept of flux balancing is solid, and it is only for very short transients that the metabolite levels within the cells will change over time.
- applying an objective function of optimizing growth is consistent with the extensive evolution of living cells to maximize their performance.

Despite the advantages of metabolic models, these models do have some shortcomings, of which the most important is the lack of regulatory features. Thus, the models can not describe conditions that heavily rely on regulation of specific parts of the metabolic network. However, as metabolic models rely on simple mass balancing of metabolites and there are no fitted parameters, they are open-ended models, and new features can therefore easily be incorporated. Hence, these models can be continuously improved in the iterative cycle of model construction as illustrated in Figure 2-1.

A very important step in generating metabolic models is the reconstruction of the metabolic network. Particularly it is important that the reconstruction process is done such that as much information as possible is build into the model. Hence the reconstruction involves several steps, and the process of genome-scale metabolic model reconstruction is illustrated in Figure 2-2. It usually starts with assigning reactions to the annotated genes in the genome sequence. Further the presence of various pathways in the cell is validated and the pathways are thoroughly curated using biochemical literature and publications. The next step is to identify the gaps in the metabolic network and try to fill them using directed search for the missing genes. The model can be validated against physiological data, 'omic' data, etc. as well as these data can be analyzed using the model. Addition of regulatory constrains improves model performance as it resolves the inconsistencies when regulation occurs on higher levels than fluxes level. Some parts of the reconstruction process (as the template model building, pathway presence evaluation, holes filling) can be automated using available software (e.g. Pathway Tools [12]), however scientific curation is still preferred for the final decisions.

Figure 2-2. The process of metabolic model reconstruction.
Information flow
Actions



A number of stoichiometric reconstructions of metabolic networks have appeared representing all three domains of life: archaea, bacteria, and eukarya. The sizes of these reconstructions range from tens to hundreds of reactions and genes; genome-scale reconstructions are beginning to become more prevalent, which is enhanced by the recent abundance of sequencing data. These genome-scale reconstructions include reactions in central metabolic pathways, such as glycolysis, pentose phosphate pathway, and citric acid cycle, as well as reactions involving amino acid and nucleotide biosynthesis, cell wall synthesis, cofactor biosynthesis, and alternate carbon and nitrogen source utilization. Table 2-2 shows the properties of various genome-scale reconstructions for different organisms. The largest reconstructions to date are for *Saccharomyces cerevisiae* and *E. coli*, and for *E. coli* these reconstructions have been used to study this model organism for over a decade [13].

Table 2-2. Total organism ORFs, number of metabolic genes, number of reactions, number of metabolites.

Organism	ORFs	Genes	Metabolites	Reactions	Reference
Escherichia coli	4,405	904	625	931	[13]
Saccharomyces cerevisiae	6,183	750	646	1149	[14]
Haemophilus influenzae	1,775	296	343	488	[23]
Helicobacter pylori	1,632	291	340	388	[24]
Plasmodium falciparum	5,342	737[a]	525	697	[25]
Mannheimia succiniproducens	2,463	335	352	373	[26]

Methanococcus jannaschii	1,821	436[a]	510	609	[27]
Streptomyces coelicolor	8,042	769	501	700	[30]
Staphylococcus aureus	2,702	619	571	641	pers. comm. B.O. Palsson
Mus musculus			872	1220	[28]

[a] only the number of enzymes were reported.

2.4 Metabolic models: a base for systems biology

Metabolic networks provide a framework for the integrated analysis of different types of high-throughput data in this way serving the purpose of systems biology – understanding the cellular behavior as a whole. In the following section we discuss the recent achievements in the area – application of metabolic models for analysis of phenotype, gene expression data, metabolites concentrations as well as for analysis of the combination of several data types.

2.4.1 Phenotype

Genome-scale metabolic models have shown to be very useful for analysis of single deletion mutants at a large scale, and a couple of studies have appeared recently for analyzing mutant phenotypes for *S. cerevisiae* and *E. coli*. The *S. cerevisiae* study looked at 4,154 cases of different mutant strains growth phenotypes in different media compositions [14]. The metabolic network reconstruction for this organism was used to accurately predict or explain 82.6% of the cases. A combined metabolic and regulatory network reconstruction of *E. coli* was used to predict growth on various carbon and nitrogen sources for a range of metabolic and regulatory gene deletions [15]. An important development in this model was addition of regulatory constraints collected from the literature, which were incorporated as Boolean logic statements. A disadvantage of using Boolean logic statements for simulating regulation is the "switch" mechanism provided for signal transduction, compared with the more likely sigmoid function (or other type of shape) for activating the individual components in a real signal transduction pathway. The predictions obtained from using Boolean logic statements were, however, in good agreement with experimental data from high-throughput phenotypic arrays. The stoichiometric network alone could correctly predict the phenotype in 65.2% of 13,750 cases, while the addition of regulatory effects improved the prediction rate to 78.7%.

2.4.2 Gene expression

The generation of gene expression data for a wide range of organisms and experimental conditions, provides a ready source of data for analysis within the context of network reconstructions. Metabolic maps are widely used for visualization of expression data, putting the data into biochemical context for analysis. The expression data can for example be visualized by color gradients on metabolic maps (e.g. Pathway Tools Omics Viewer [12]). A number of methods that rely on reconstructed metabolic networks have been developed that provide a more rigorous way of analyzing gene expression data.

Ihmels et al. [16] used the reaction set for *Saccharomyces cerevisiae* derived from Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene expression data for over 1,000 conditions from public repository for identifying correlations between expressions of the genes from the same metabolic pathway using signature

algorithms. They observed that enzymes involved in the same metabolic pathway were often co-expressed, and that the co-expressed enzymes were usually arranged in a linear pathway. The authors also found that at the branch points, the incoming branch is normally co-regulated with only one of the outcoming branches, which leads to linear metabolic flow through the network.
Ihmels' et al. work was based only on the connectivity between substrates and did not consider molecules involved in energy metabolism as well as reducing cofactors, which are known to be the most connected metabolites in the networks [17]. Incorporation of these compounds in the analysis might reveal less obvious enzyme interactions mediated by these common metabolites.

Patil and Nielsen [18] applied a genome-scale metabolic model of *S. cerevisiae* [19] for the transcriptome analysis of several mutants in comparison to the wild type and for analysis of expression data from wild type strains grown on several different carbon sources. This computational approach was based on a graph-theoretical representation of the metabolic network and allowed identification of sub-networks with significantly correlated changes in gene expression as well as metabolites around which the most significant expression changes occurred. Because finding of high scoring sub-networks is a computationally extensive task, it was solved by using a simulated annealing algorithm.

An analogical approach was reported previously for identifying highly scoring sub-networks in a molecular interaction network [20]. In this study expression data was used from yeast mutants in the galactose assimilation pathway grown with and without galactose in combination with protein-protein and protein-DNA interactions obtained from publicly available databases. Many of the identified highly scoring sub-networks coincided with the previously known regulatory circuits.

The previously mentioned combined metabolic and regulatory networks reconstructed for *E. coli* were also used to demonstrate the iterative process of regulatory network reconstructions through the analysis of gene expression data [15]. The genome-scale network was used to generate predictions about gene expression in response to oxygen deprivation. When these predictions were inconsistent with the gene expression data, the regulatory constrains were modified. One can however discuss how suitable the prediction methodology is because in genome-scale networks multiple solutions normally exist. Besides taking into consideration that the initial regulatory network was obtained from the literature, the change of regulatory rules should only be viewed as a testable hypothesis about the involvement of different transcription factors in the regulation of metabolic genes.

2.4.3 Metabolites concentrations

Given that metabolic networks can be reconstructed at a genome-scale, an obvious data type that could be accounted for would be metabolite concentrations themselves. To date, intracellular metabolite concentration measurements have been taken primarily for plants. While the use of networks for analysis of such data is not as common as gene-expression data, the potential of such data will likely lead to the development of methods for analyzing it.

A significant progress towards this goal was made by Weckwerth et al. in a study of silent potato plant phenotype using metabolomics [21]. Over 1,000 compounds were quantitatively measured by GC/TOF MS in more than 30 samples for the wild type and mutant potato lines. The authors identified strong linear correlations (Pearson correlation coefficient more than 0.8) between metabolites concentrations and used these pair-wise correlations to represent the network topology for the wild type and mutant. The analysis of metabolites connectivity in the obtained networks indicated changes in carbohydrate and amino acids metabolism caused by silent mutation.

2.4.4 Gene expression, proteome, interactome, phenotype

Ideker *et al.* [22] used an integrated approach to study a pathway of galactose utilization (also see update). They started with reconstructing a network including enzymatic and transport reactions as well as protein-protein and protein-DNA interactions. The model was then used to make predictions about the behavior of the GAL mutants grown with and without galactose. Apparently the predictions were made *ad hoc*, which undermines the modeling aspect of this work and the applicability of the method for larger and/or less studied systems. The effects of the perturbation of the pathway by genetic modifications and environmental conditions were evaluated by transcriptome and proteome analyses. The work heavily relied on the visualization of the data. The genes with significantly changed expression profiles were overlaid on the physical interaction map and turned out to be highly connected. Clustering analysis of transcription data for several mutants strains allowed to confirm some initial hypothesis about the gene functions.

2.5 Perspective

The integrated analysis of 'omic' data is in its germination stage. However already the first results demonstrate how powerful the techniques can be and how much more information can be extracted from the same data when the analysis is supplemented with metabolic or (better and) regulatory networks. There's a need for high quality metabolic and regulatory networks, which should preferably be presented in the same format. A challenging task is ahead of finding new computational and visualization methods, that integrate different data types. The tools should be presented in a user friendly format to allow numerous researches to gain more information from their experiments. In this context the development of systems biology markup language (SBML) is important. The field of systems biology will also benefit from having high quality network reconstructions. It is important that these reconstructions represent the biological system of interest, as such there needs to be a balance between being able to rapidly generate these reconstructions and to curate them. This curation can arise from the available literature for an organism as well as feedback or iterative improvements of the network reconstruction via analysis of high throughput data.

2.6 Update

In a recent study the impact of arsenic in *S. cerevisiae* was studied by mapping the gene expression and phenotypic sensitivity data onto metabolic and regulatory networks [29].

2.7 References

1. Westerhoff HV, Palsson BO: **The evolution of molecular biology into systems biology.** *Nat Biotechnol* 2004, 22:1249-1252.
2. de Lichtenberg U, Jensen LJ, Fausboll A, Jensen TS, Bork P, Brunak S: **Comparison of computational methods for the identification of cell cycle regulated genes.** *Bioinformatics* 2004.
3. de Lichtenberg U, Jensen LJ, Brunak S, Bork P: **Dynamic complex formation during the yeast cell cycle. *Science* 2005, in press.
The authors integrate microarray time-series data with the information on protein-protein interactions to study periodically expressed genes in the yeast cell cycle.
4. Pritchard L, Kell DB: **Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis.** *Eur J Biochem* 2002, 269:3894-3904.
5. Hoefnagel MH, Starrenburg MJ, Martens DE, Hugenholtz J, Kleerebezem M, Van S, II, Bongers R, Westerhoff HV, Snoep JL: **Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis.** *Microbiology* 2002, 148:1003-1013.
6. Price ND, Papin JA, Schilling CH, Palsson BO: **Genome-scale microbial in silico models: the constraints-based approach.** *Trends Biotechnol* 2003, 21:162-169.
7. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110.** *Appl Environ Microbiol* 1994, 60:3724-3731.
8. Palsson B: **Two-dimensional annotation of genomes.** *Nat Biotechnol* 2004, 22:1218-1219.
9. Edwards JS, Palsson BO: **The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities.** *Proceedings of the National Academy of Sciences* 2000, 97:5528-5533.
10. Forster J, Famili I, Palsson BO, Nielsen J: **Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*.** *Omics* 2003, 7:193-202.
11. Ibarra RU, Edwards JS, Palsson BO: ***Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* 2002, 420:186-189.
The authors show, that under the growth selection pressure *E. coli* optimizes its growth on glycerol to achieve the optimal value predicted from the genome-scale stoichiometric model.
12. Karp PD, Paley S, Romero P: **The Pathway Tools software.** *Bioinformatics* 2002, 18:225-232.
13. Reed JL, Palsson BO: **Thirteen Years of Building Constraint-Based In Silico Models of *Escherichia coli*.** *J Bacteriol* 2003, 185:2692-2699.
14. Duarte NC, Herrgard MJ, Palsson B: **Reconstruction and Validation of *Saccharomyces cerevisiae* iND750, a Fully Compartmentalized Genome-Scale Metabolic Model.** *Genome Res* 2004, 14:1298-1309.
*15. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO: **Integrating high-throughput and computational data elucidates bacterial networks.** *Nature* 2004, 429:92-96.
This paper is the first attempt to incorporate regulatory rules into a stoichiometric genome-scale model. Addition of regulation into *E. coli* model improves the phenotype predictions for the large-scale phenotype arrays data. The model is also used to calculate the outcome of gene expression data under oxygen

limitation and to modify the regulatory constraints when the model leads to wrong predictions.

16. Ihmels J, Levy R, Barkai N: **Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat Biotechnol* 2004, 22:86-92.

The authors use large-scale gene expression data (over 1,000 profiles) along with the structure of the metabolic network derived from KEGG database to study how pathways are regulated on transcriptional level.

17. Nielsen J: **It is all about metabolic fluxes.** *J Bacteriol* 2003, 185:7031-7035.

18. Patil K, Nielsen J: **Uncovering transcriptional regulation of metabolism using metabolic network topology. *Proc Natl Acad Sci U S A* 2005, 102:2685-2689.

In this work gene expression data is analyzed in the context of the topology of *S. cerevisiae* genome-scale metabolic network. This allows identification of transciptionally coupled metabolic genes sets that are not detectable by traditional statistical microarray data analyses.

19. Forster J, Famili I, Fu PC, Palsson BO, Nielsen J: **Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network.** *Genome Res* 2003, 13:244-253.

20. Ideker T, Ozier O, Schwikowski B, Siegel AF: **Discovering regulatory and signalling circuits in molecular interaction networks.** *Bioinformatics* 2002, 18:233-240.

*21. Weckwerth W, Loureiro ME, Wenzel K, Fiehn O: **Differential metabolic networks unravel the effects of silent plant phenotypes.** *Proc Natl Acad Sci U S A* 2004, 101:7809-7814.

The linear correlations between metabolite concentrations are used to construct metabolite interactions maps for the wild type potato plant and its antisense mutant with a silent phenotype. Comparison of the topologies of these maps leads to hypotheses about the influence of the mutation on the cellular metabolism.

22. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science* 2001, 292:929-934.

23. Edwards JS, Palsson BO: **Systems properties of the *Haemophilus influenzae* Rd metabolic genotype.** *Journal of Biological Chemistry* 1999, 274:17410-17416.

24. Schilling CH, Palsson BO: **Assessment of the Metabolic Capabilities of *Haemophilus influenzae* Rd through a Genome-scale Pathway Analysis.** *Journal of Theoretical Biology* 2000, 203:249-283.

25. Yeh I, Hanekamp T, Tsoka S, Karp PD, Altman RB: **Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery.** *Genome Res* 2004, 14:917-924.

26. Hong SH, Kim JS, Lee SY, In YH, Choi SS, Rih JK, Kim CH, Jeong H, Hur CG, Kim JJ: **The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*.** *Nat Biotechnol* 2004, 22:1275-1281.

27. Tsoka S, Simon D, Ouzounis CA: **Automated metabolic reconstruction for *Methanococcus jannaschii*.** *Archaea* 2004, 1:223-229.

28. Sheikh K, Forster J, Nielsen LK: **Modeling Hybridoma Cell Metabolism Using a Generic Genome-Scale Metabolic Model of *Mus musculus*.** *Biotechnol Prog* 2005, in press.

29. Haugen AC, Kelley R, Collins JB, Tucker CJ, Deng C, Afshari CA, Brown JM, Ideker T, Van Houten B: **Integrating phenotypic and expression profiles to map arsenic-response networks.** Genome Biol 2004, 5:R95.
30. Borodina I, Krabben P, Nielsen J: **Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism.** Genome Res 2005, in press.

CHAPTER 3

Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism

Irina Borodina[1], Preben Krabben[2], and Jens Nielsen[1]

[1] Center for Microbial Biotechnology, BioCentrum-DTU, Building 223, DK-2800 Kgs. Lyngby, Denmark
2 Department of Biochemical Engineering, University College London, Torrington Place, London, UK

Personal communication with:
Stephen Bentley (sdb@sanger.ac.uk), The Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

Content of this chapter is published in: Genome Research, 2005, 15(6):820-829.

3.1 Abstract

Streptomyces are filamentous soil bacteria that produce more than half of the known microbial antibiotics. We present the first genome-scale metabolic model of a representative of this group – *S. coelicolor* A3(2). The metabolism reconstruction was based on annotated genes, physiological and biochemical information. The stoichiometric model includes 819 biochemical conversions and 152 transport reactions, accounting for a total of 971 reactions. 700 of the reactions in the network are unique, while the rest are iso-reactions. The network comprises 500 metabolites. Seven hundred eleven open reading frames (ORFs) were included in the model, which corresponds to 13% of the ORFs with assigned function in the *S. coelicolor* A3(2) genome. In a comparative analysis with the *S. avermitilis* genome we showed that the metabolic genes are highly conserved between these species and therefore the model is suitable for use with other Streptomycetes. Flux balance analysis was applied for studies of the reconstructed metabolic network and to assess its metabolic capabilities for growth and polyketides production. The model predictions of wild type and mutants' growth on different carbon and nitrogen sources agreed with the experimental data in most cases. We estimated the impact of each reaction knock-out on the growth of *in silico* strain on 62 carbon sources and 2 nitrogen sources; thereby identifying the "core" of the essential reactions. We also illustrated how reconstruction of a metabolic network at the genome level can be used to fill gaps in genome annotation.
The supplemental material is available on www.genome.org

Keywords

Streptomyces coelicolor, *in silico*, metabolic model.

3.2 Introduction

Streptomycetes are soil inhabiting Gram-positive bacteria belonging to the order of *Actinomycetales* (Garrity 2002). The genus is characterized by a high G+C content, large linear chromosomes (>8 Mb) and a complex life cycle, which involves morphological differentiation. Streptomycetes have attracted much attention due to their ability to make secondary metabolites that have a wide range of bio-activities and hereby may find use as antibiotics, immunosuppressants, anti-cancer agents, etc. It is estimated that *Streptomyces* spp. produce more than 50% of the total of 11,900 known microbial antibiotics (Kieser et al. 2000). *Streptomyces* species also secrete a wide range of extracellular hydrolytic enzymes, which enable them to degrade lignocellulosic materials in soil, and many of these enzymes also have commercial interest.

Streptomyces coelicolor A3(2) is by far the best genetically studied *Streptomyces* strain and has become a model organism for *Streptomyces* species (Hopwood 1999). Release of the *S. coelicolor* A3(2) genome sequence (Bentley et al. 2002) has further expanded the knowledge of this organism and enabled large-scale analyses of transcriptome (Huang et al. 2001; Bucca et al. 2003) and proteome (Hesketh et al. 2002; Novotna et al. 2003). The rapidly growing amount of genome-wide data represents a valuable database for gaining more fundamental insight into the molecular mechanisms governing different processes in *S. coelicolor* A3(2), but particularly integration of these data through the use of mathematical models would enable extraction of more information from this database. In this context it has

recently been shown particularly valuable to use genome-scale models for the metabolism (Ideker et al. 2001; Stelling et al. 2002; Kell 2004; Covert et al. 2004).
At the present, metabolic networks have been reconstructed for several bacteria (e.g. *Escherichia coli* (Edwards et al. 2000b; Reed et al. 2003),) and for the yeast *S. cerevisiae* (Förster et al. 2003a; Duarte et al. 2004). The models have been successfully used to predict organism phenotype for a modified genotype (Edwards et al. 2000a; Förster et al. 2003b; Covert et al. 2004), and to predict cellular behavior under different physiological conditions (Edwards et al. 2001; Famili et al. 2003). Furthermore, these models may form the basis for the functional characterization (Papp et al. 2004), for extraction of information of coordinated transcriptional responses that can not be identified through classical statistical methods (Ihmels et al. 2004), and in design of improved strains for production of specific metabolites (Burgard et al. 2003).
Here, the reconstructed metabolic network of *S. coelicolor* A3(2) is applied for detailed study of the organism's metabolism.

3.3 Results and discussion

3.3.1 Reconstruction and characteristics of the metabolic network

Reconstruction of high-quality and well annotated metabolic networks is laborious as information needs to be acquired from many different sources. The reconstruction can be facilitated by software, e.g., PathoLogic (Karp et al. 2002), that assigns reactions to the annotated genes according to E.C. number or enzyme name. We used *S. coelicolor* A3(2) pathway database in KEGG (ftp://ftp.genome.ad.jp/pub/kegg/pathways/sco/) for extraction of the draft model. Automatically created models need to be manually curated using books, literature and other available information sources. The most problematic aspects are wrongly or insufficiently defined substrate specificity, reactions reversibility, protein complexes, cofactor specificity and the missing enzymes. Examples of these problems and the ways of dealing with them are listed in Table 3-1.

Table 3-1. Typical problems arising from automatic metabolic networks reconstruction and examples of dealing with them.

Problem	Example
1) Substrate specificity is insufficiently or wrongly defined.	The ORF SCO4384 encodes an enoyl-CoA hydratase with assigned E.C. number 4.2.1.17 and the enzyme acts on a wide range of 3-hydroxyacyl-CoA compounds. Judging from the neighbor ORF SCO4383, which encodes a 4-coumarate:coenzyme A ligase and is positioned in the same regulon, SCO4384 more likely catalyzes hydration of activated 4-coumarate and hence is involved in the secondary metabolism.
2) Reversibility is not defined.	Pyruvate carboxylase (E.C. 6.4.1.1) is shown in KEGG as acting in both directions even though it is known to be irreversible due to thermodynamic constraints prevailing in most cells, i.e. the carboxylation of pyruvate to oxaloacatate is associated with the cleavage of a high-energy phosphate bond in ATP resulting in the formation of AMP and pyrophosphate whereas the decarboxylation of oxaloacetate to pyruvate does not result in the formation of ATP.
3) Enzyme subunits are shown as catalyzing a reaction independently though they are active only in a complex.	The ORFs SCO0216-0219 form a four-subunit complex with a nitrate reductase activity, while in KEGG each of these ORFs is assigned the given function, basically indicating that each of these four ORFs encodes an isoenzyme.
4) Cofactor requirements are often specific for the given	Valine dehydrogenase E.C. 1.4.1.8 is given as $NADP^+$-dependent in KEGG, however it was experimentally determined to use NAD^+ as the

organism and have to be found elsewhere.	preferred hydrogen acceptor (Navarrete et al. 1990).
5) A number of reactions that are necessary for making a functional cell have not been assigned a corresponding ORF.	Histidinol phosphatase E.C. 3.1.3.15 is an enzyme in the linear histidine biosynthesis pathway. It was included in the model even though the gene was not found in the genome of *S. coelicolor* A3(2).

Modeling of *S. coelicolor* A3(2) is a challenging task due to the complexity of this filamentous bacteria. Its genome size of 8,667,507 base pairs is unusually large for bacteria sequenced to date. Out of 7,825 predicted open reading frames (ORFs), a function has been assigned to 5,492, which also includes broad definitions as "putative membrane protein", "putative transmembrane transport protein" and similar (personal communication with Bentley S.D.). It is estimated that 965 proteins have a regulatory function, 819 proteins are secreted (these include numerous hydrolytic enzymes), 614 proteins participate in transport and twenty-two gene clusters are involved in secondary metabolites production (Bentley et al. 2002). Only 926 ORFs have an enzyme commission (E.C.) number assigned and out of these, 711 (77 %) were included in the model (Table 3-2, Appendix 3.1).

Table 3-2. Comparison of the genome characteristics and the *in silico* metabolic networks of prokaryotic and eukaryotic species.

	Prokaryotes				Eukaryotes
	***S. coelicolor* A3(2)**	***E. coli*[1]**	***H. influenzae*[2]**	***H. pylori*[3]**	***S. cerevisiae*[4]**
Genome characteristics					
Genome length (bp)	8,667,507	4,639,221	1,830,135	1,667,867	12,175,026
G+C content	72%	51%	38%	39%	40%
ORF	7,825	4,288	1,743	1,590	6,281
assigned function	5,492	2,656	1,011	1,091	4,379
no assigned function	2,333	1,632	732	499	1,902
***In silico* metabolic network characteristics**					
Total reactions (unique)	971 (700)	931 (931)	461	381	1175 (842)
biochemical conversions	819	767			826
Transport	152	164			349
Reactions with ORF (% of total reactions)	764 (79%)	873 (94%)	412 (89%)	272 (71%)	1035 (88%)
Metabolites	500	625	367	332	584
ORF (% of ORFs with assigned function)	711 (13%)	904 (34%)	400 (40%)	290 (27%)	708 (16%)

[1] *Escherichia coli* K-12 (Reed et al. 2003)
[2] *Haemophilus influenzae* Rd (Schilling et al. 2000)
[3] *Helicobacter pylori* 26695 (Schilling et al. 2002)
[4] *Saccharomyces cerevisiae* (Förster et al. 2003a)

We considered the pathways, which were known to be operative in *Streptomyces* and/or for which most of the pathway specific enzymes were present. Genome-scale modeling is a dynamic process and we expect that the reaction list will be updated as new information is released. The total number of metabolic reactions included in the model is 971, of which 819 are metabolic conversions and the rest are transport reactions (Appendix 3.2). 89% of the metabolic conversion reactions have an ORF assigned; the other reactions were included based on physiological evidence. Only 24% of the transport reactions have an assigned ORF, and this low percentage is due to the limited information on transport processes in this organism. The model

includes biosynthesis of the polyketide antibiotic actinorhodin (ACT) and a non-ribosomal peptide calcium-dependent antibiotic (CDA). The biosynthesis route of undecylprodigiosin (RED) is not yet completely elucidated (Thomas et al. 2002), therefore only the known steps for its biosynthesis are presented in the model. With the included reactions the total number of metabolites in the reconstructed network is 500 (Appendix 3.3).

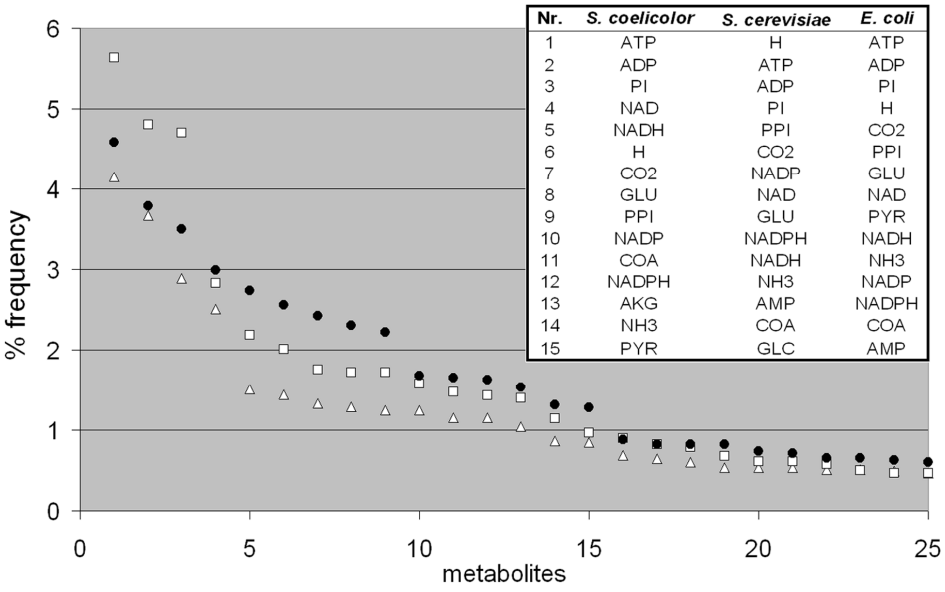
3.3.2 Universality of the S. coelicolor A3(2) model

We estimated the possibility of extrapolating the use of the *S. coelicolor* A3(2) metabolic network for other *Streptomyces* strains by comparing the genomes of *S. coelicolor* A3(2) and *S. avermitilis* (Appendix 3.4). In total 53% of the genes in *S. coelicolor* A3(2) are synteny conserved with *S. avermitilis*, i.e. their relative position in the genome has been retained throughout evolution of these two *Streptomyces* strains; 35% of the *S. coelicolor* A3(2) genes have orthologs, which position is not conserved in *S. avermitilis*; only 12% of the genes do not have an ortholog in *S. avermitilis*. Out of 711 ORFs included in the model 78% are synteny conserved with *S. avermitilis*, which shows that there is a much higher degree of synteny for metabolic genes than for the other genes. A further 22% of the ORF's in the metabolic model are coding for isoenzymes of which at least 1 synteny conserved ORF is present in *S. avermitilis*. The high fraction of synteny conserved ORFs and isoenzymes in the model indicates that the *S. coelicolor* A3(2) metabolic network can be used as a starting point for a rapid reconstruction of the reaction networks for other *Streptomyces* species.

3.3.3 Connectivity

A frequency plot for all the metabolites in the metabolic networks of *S. coelicolor* A3(2), *Saccharomyces cerevisiae* (Förster et al. 2003a) and *E. coli* (Edwards et al. 2000b) is shown in Figure 3-1. With the exception of proton, which is the most frequent metabolite in yeast due to proton-driven transport reactions, the most connected metabolites are involved in energy metabolism: ATP, ADP and phosphate. The reducing equivalent NADH is more common for the metabolic network of *S. coelicolor* A3(2), which also has the highest proportion of oxidoreductases. The higher occurrence of glutamate and α-ketoglutarate in *S. coelicolor* A3(2) reflects a higher frequency of the use of aminotransferases. Coenzyme A and acetyl-carrier protein occur more often in the *S. coelicolor* A3(2) model, which directly correlates with extensive propanoate, butanoate metabolism and the production of polyketide actinorhodin.

Figure 3-1. Frequency of the most connected metabolites in the reconstructed metabolic networks of *S. coelicolor* A3(2), *Saccharomyces cerevisiae* (Förster et al. 2003a) and *E. coli* (Edwards et al. 2000b). Frequency is calculated as the number of times a certain metabolite appears in the metabolic network divided by the sum of all metabolites occurrence. *S. coelicolor* A3(2) (●), *Saccharomyces cerevisiae* (∆), *E. coli* (□). Abbreviations: H – external proton, GLU – glutamate, PPI – pyrophosphate, COA – coenzyme A, AKG – α-ketoglutarate, PYR – pyruvate, ACCOA – acetyl-coenzyme A, ACP – acetyl-carrier protein.



Nr.	*S. coelicolor*	*S. cerevisiae*	*E. coli*
1	ATP	H	ATP
2	ADP	ATP	ADP
3	PI	ADP	PI
4	NAD	PI	H
5	NADH	PPI	CO2
6	H	CO2	PPI
7	CO2	NADP	GLU
8	GLU	NAD	NAD
9	PPI	GLU	PYR
10	NADP	NADPH	NADH
11	COA	NADH	NH3
12	NADPH	NH3	NADP
13	AKG	AMP	NADPH
14	NH3	COA	COA
15	PYR	GLC	AMP

3.3.4 Growth energetics

Besides differences in some reactions and the preferred cofactors for some reactions, the core part of microbial metabolism, i.e. the central carbon metabolism (glycolysis, pentose phosphate pathway, and TCA cycle), biosynthesis of the biomass building blocks, is highly conserved. The biomass composition and the growth energetics are, however, unique properties of each organism. Here, the biomass composition was calculated based on the measurements performed in *S. coelicolor* A3(2) and related species (Appendix 3.5). The growth energetics is described by Stouthamer equation (1):

$$r_{ATP} = Y_{xATP} \cdot \mu + m_{ATP}, \qquad (1)$$

which states that for a pseudo-steady state the rate of ATP production r_{ATP} is equal to the ATP consumption for growth $Y_{xATP} \cdot \mu$ and maintenance m_{ATP}, where μ is the specific growth rate and Y_{xATP} is the ATP yield coefficient (Stouthamer et al. 1973). The main source of ATP in aerobically growing organisms is the oxidative phosphorylation. *Streptomyces* are known to use menaquinone as electron transporter in the membrane (Pandya et al. 1966; Yassin et al. 1988), even though surprisingly enough the genes for menaquinone synthesis were not found in the sequenced genomes of *Streptomyces coelicolor* A3(2) and *Streptomyces avermitilis*. The stoichiometry of the electron transport chain was assumed to be the same as recently described in the actinomycete *Corynebacterium glutamicum* (Bott et al. 2003). The values of the ATP yield coefficient Y_{xATP} and the maintenance flux m_{ATP} were estimated from the chemostate cultivations data (Melzoch et al. 1997) as described in Methods (Figure 3-2, 3-3).

Figure 3-2. The ATP yield coefficient Y_{xATP} (○) and maintenance energy m_{ATP} (■) as a function of the maximal P/O ratio in *Streptomyces coelicolor* A3(2).

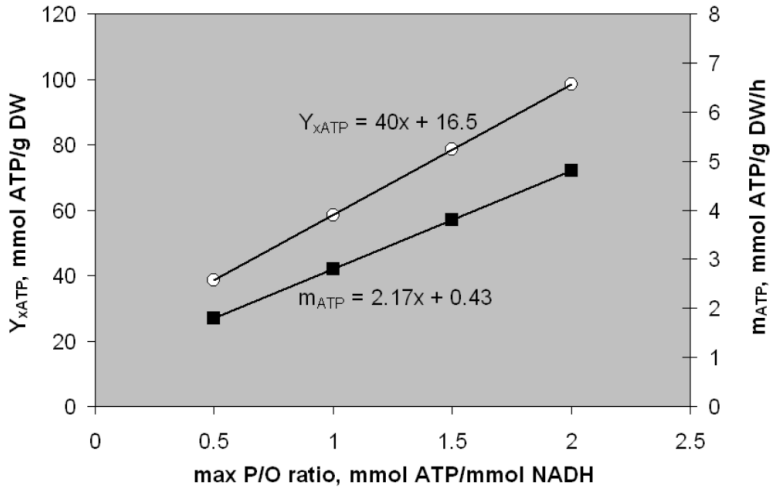
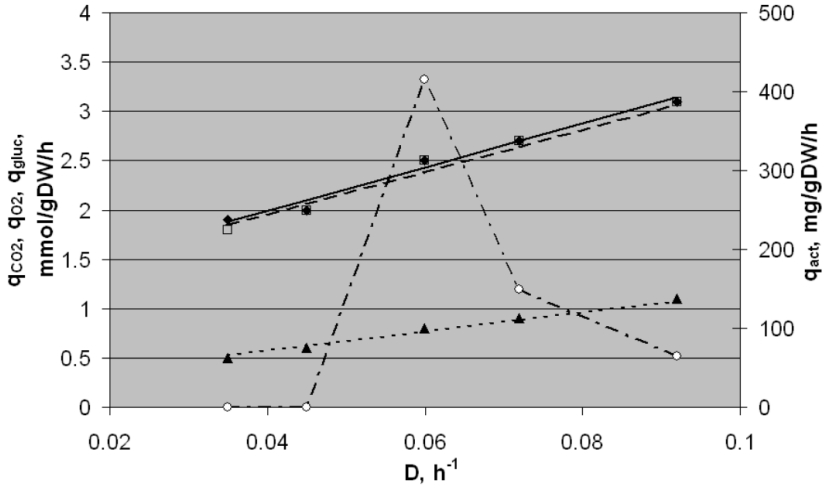


Figure 3-3. Simulation of experimental chemostates data (Melzoch et al. 1997). Specific glucose uptake rate (experimental ▲, model ----), specific carbon dioxide production rate (experimental ♦, model ——), specific oxygen uptake rate (experimental □, model – – –), actinorhodin production rate (experimental o, model — – — –).

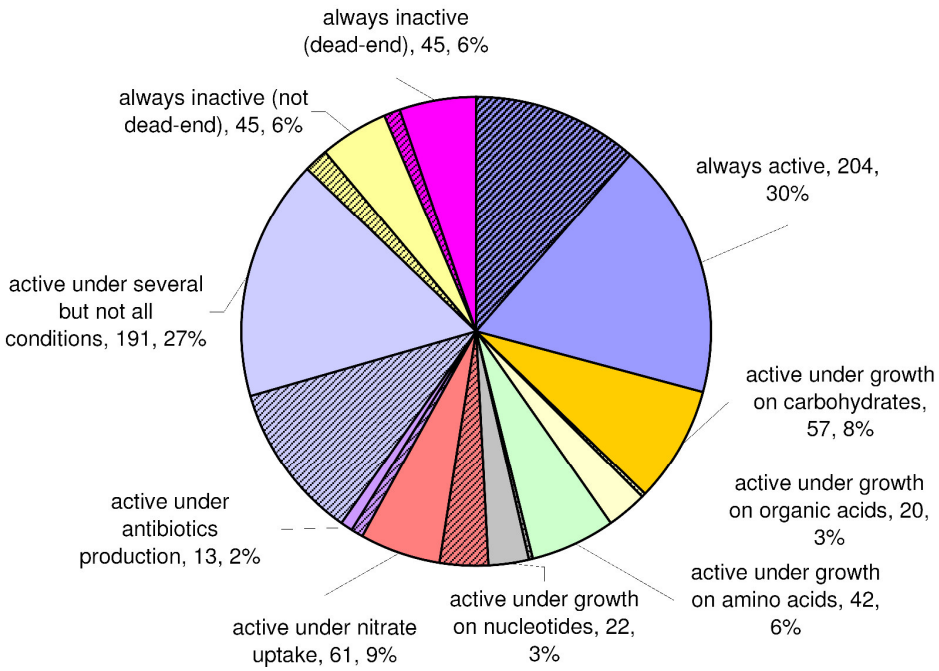


3.3.5 Reactions activity

In order to investigate the topology of the metabolic network we created a list of unique reactions, and in this way eliminated the redundancy in the network resulting from the large number of isoenzymes (Appendix 3.7). This model was used for growth simulation in five different media: 1) glucose as C-source and nitrate as N-source, 2) mixture of carbohydrates, 3) mixture of organic acids and alcohols, 4) mixture of amino acids, and 5) mixture of nucleotides (Appendix 3.8). Except for the first scenario, non-limited ammonia uptake was allowed. Additionally, we simulated secondary metabolites production at slow growth rate on glucose and ammonia. The

analysis of the fluxes that computationally provided optimal growth rate or optimal antibiotics production rate at the given conditions (also considering alternative optimal solutions) showed that 88% of reactions were active under some or all of the tested conditions, 6% of reactions were inactive due to the formation of dead-ends (that means that one of reaction's products or reactants can not be used or produced in the network (Figure 3-4)). In this way there was only a very small part (6%) of the non dead-end reactions that never got involved in the cellular metabolism, perhaps because the conditions that evoke these reactions' activity have not been considered in the simulations.

Figure 3-4. Activity of the reactions in the model of *S. coelicolor* A3(2). All the unique reactions in the model are divided into a number of categories based on their predicted activity under several growth conditions. The proportion of reactions for which the corresponding enzyme (-s) were detected on a 2D-gel is shown by dashed area.



Interestingly, some of the enzymes that catalyze "dead-end" reactions in the model have been detected on 2D gels in *S. coelicolor* A3(2) grown on a minimal medium supplemented with casamino acids (http://dbkweb.ch.umist.ac.uk/StreptoBASE/s_coeli/referencegel/). One of the enzymes was putative mannose-1-phosphate guanyltransferase (SCO1388), which is known to be involved in mannosylation of proteins, a common protein modification for Actinomycetes. The reaction appears as a dead-end in the model because protein modifications are not included in the reactions list and therefore the GDP-activated mannose-1-phosphate is not used in any other reactions. Another enzyme is 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase (SCO0852), which is either involved in Entner-Doudoroff (ED) pathway (this is unlikely as the characteristic ED pathway gene *edd* was not found in the genome sequence of *Streptomyces coelicolor* A3(2)) or is responsible for interconverting 4-

hydroxy-α-ketoglutarate into glyoxylate and pyruvate. Experimental evidence has been obtained that the enzyme is important in regulation of glyoxylate levels in the cells of *E. coli* (Cayrol et al. 1995). A hypothesis has been proposed that 4-hydroxy-2-oxoglutarate aldolase action is coupled to α-ketoglutarate dehydrogenase complex and results in pyruvate-catalyzed oxidation of glyoxylate into malyl-CoA (Gupta et al. 1984). We judge that it is important to include the reactions in the model even though the complete pathway is not known as dead-ends presence may give a hint to which enzymes one should look for in future annotation of the genome.

3.3.6 Reactions dispensability

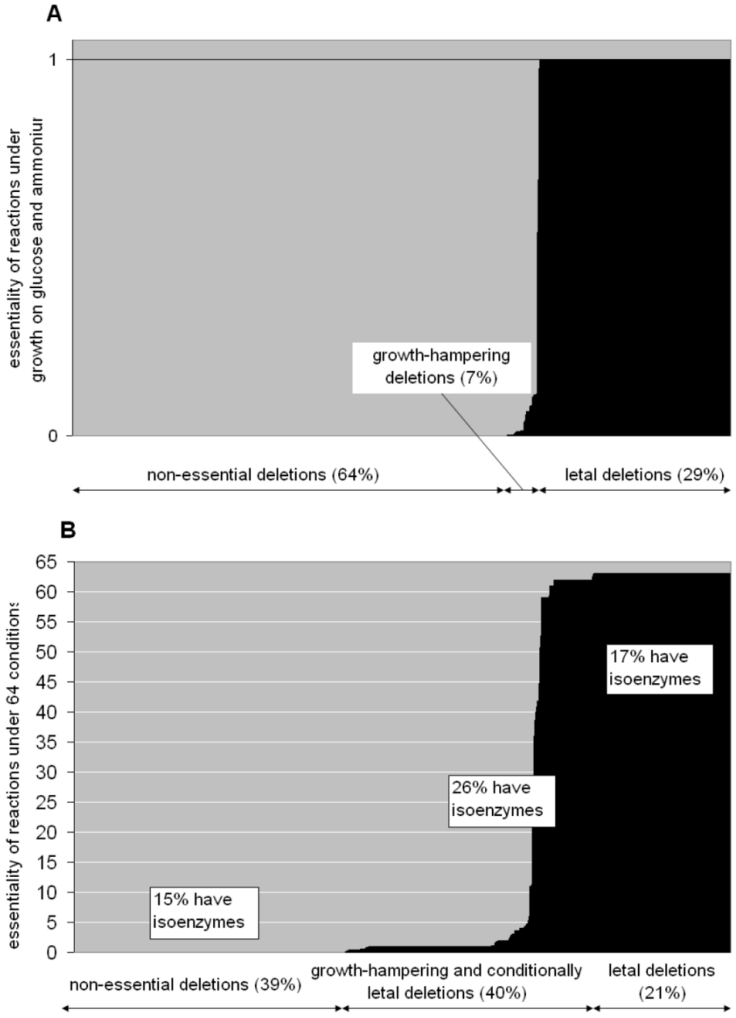
We studied dispensability of the reactions in the network by making single reaction deletions and optimizing for the growth on 62 different carbon sources and 2 different inorganic nitrogen sources (ammonia and nitrate, only tested with glucose as carbon source). The approach differs from single gene deletion studies, because if isoenzymes are present, then their simultaneous knock-out will be simulated by deleting the reaction they catalyze. We find this approach more informative for studying the sensitivity of the metabolic network to perturbations. To quantify the effect of reaction deletion we defined reaction essentiality as the relative decrease in the specific growth rate with deletion of the reaction in comparison to the specific growth rate with the complete reaction set (2):

$$essentiality_{reaction\ i} = 1 - \frac{\mu_{\max}(mutant_{\Delta reaction\ i})}{\mu_{\max}(wild\ type)} = [0....1], \qquad (2)$$

i.e. the reaction essentiality is 1 for an essential reaction whereas it is 0 for a reaction that upon removal has no growth retarding effect. The reaction essentialities for the growth on glucose and ammonia are shown on Figure 3-5A, and Figure 3-5B illustrates the summed reactions essentialities for growth on the 63 different media (Appendix 3.9). Reactions with a summed reaction essentiality of 63 are required for growth under all the defined conditions, and hence are true essential reactions. The reactions that have a summed reaction essentiality lower than 63 are either necessary only under certain conditions or their deletion leads to growth retardation. The comparison of Figure 3-5A and Figure 3-5B shows that the choice of conditions is important for defining the dispensability of reactions. While during growth on glucose and ammonia 64% of the reactions could be eliminated without any consequences for the cellular growth, a number of these reactions turned out to be essential for growth on other carbon sources than glucose, and the number of non-essential deletions were reduced from 64% to 39% when all the growth conditions were considered. The minimal metabolic net, necessary for the growth under all the conditions, consists of 146 reactions. Considering that 17 % of them have an isoenzyme, there are basically 121 metabolic genes that are truly essential (e.g. essential on the complete medium), which corresponds to 12% of the original metabolic network.

Essential and non-essential reactions were found to have almost equal occurrence of isoenzymes (17 and 15% respectively), whereas a higher fraction of the growth-retarding or conditionally essential reactions were catalyzed by isoenzymes (26%). This undermines a hypothesis that isoenzymes exist to increase the robustness of the metabolic network to mutations.

Figure 3-5. Essentiality of the reactions in the model of *S. coelicolor* A3(2) during growth on glucose (A) and during growth under 63 various conditions (B). All the reactions were sorted into three categories: essential reactions, conditionally essential reactions and non-essential reactions depending on their influence on biomass synthesis. The percentage of reactions with isoenzymes among all the reactions in the given category is shown.



3.3.7 Metabolic capabilities of the network

3.3.7.1 Degradation of C- and N-sources

In most of cases the model correctly predicted growth capability on various C- and N-sources for *S. coelicolor* A3(2) wild type and mutant (Table 3-3). There was a disagreement on usage of aspartate and glutamine as the sole C-source, which was possible according to the model, but has not been observed experimentally. It is known than *S. coelicolor* A3(2) can grow on asparagine, which is presumably degraded via asparatate. In this case, it may be regulatory events that do not allow asparatate utilization rather than the lack of metabolic capacity. *S. coelicolor* A3(2) can use glutamine as the sole N-source, but not as C-source, while glutamate can be used as both. There's evidence that glutamate is decarboxylated into γ-

aminobutanoate upon the uptake (Inbar et al. 1991). Therefore, if glutamine is deaminated by glutaminase intracellularly, the formed glutamate can not further be degraded. In this way glutamine would only provide nitrogen for growth, but not carbon units.

The deletions of *trpC1* and *trpD1* in *S. coelicolor* A3(2) result in auxotrophy for tryptophan, because the genes with analogical functions – *trpC2* and *trpD2* – are located in the calcium-dependent antibiotic biosynthetic cluster and apparently can be used exclusively in the secondary metabolism (Hu et al. 1999b).
The described inconsistencies can be resolved in future by expanding the model to include regulatory constraints.

Table 3-3. Comparison of experimentally determined and *in silico* predicted growth phenotypes. + stands for growth, - for no growth.

in vivo	*in silico*	Strain	Substrate/phenotype	Literature
C-source				
+	+	*wt*	glycerol, L-arabinose, D-arabitol, D-ribose, D-xylose, L-xylitol, D-fructose, D-galactose, D-gluconate, D-glucose, D-mannitol, D-mannose, L-rhamnose, salicin, cellobiose, lactose, maltose, melibiose, trehalose, acetate, citrate, lactate, malate, pyruvate, succinate, tartrate, propanoate, alanine, asparagine, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, proline, serine, threonine, tryptophan, valine	(Hodgson 1980; Zhang et al. 1996c)
		Δ*msdA*	propanoate	(Zhang et al. 1996b)
		Δ*glnA*	no glutamine auxotrophy	(Fink et al. 1999)
-	-	*wt*	succrose, arginine, methionine, phenylalanine	(Hodgson 1980)
		Δ*msdA*	valine	(Zhang et al. 1996a)
		Δ*hisA*	trp^- his^-	(Barona-Gomez et al. 2003)
		Δ*malE*	maltose	(van Wezel et al. 1997)
		Δ*argG*	arg^-	(Flett et al. 1987)
		Δ*metF*	met^-	(Blanco et al. 1998)[1]
-	+	*wt*	aspartate, glutamine	(Hodgson 1980)
		Δ*trpC1*Δ*trpD1*	trp^-	(Hu et al. 1999a)
N-source				
+	+	*wt*	aspartate, asparagine, glutamate, glutamine, isoleucine, leucine, lysine, praline, valine	(Hodgson 1980)
-	-	Δ*vdh*	valine, isoleucine, leucine	(Tang et al. 1993)
+	-	*wt*	methionine, phenylalanine	(Hodgson 1980)

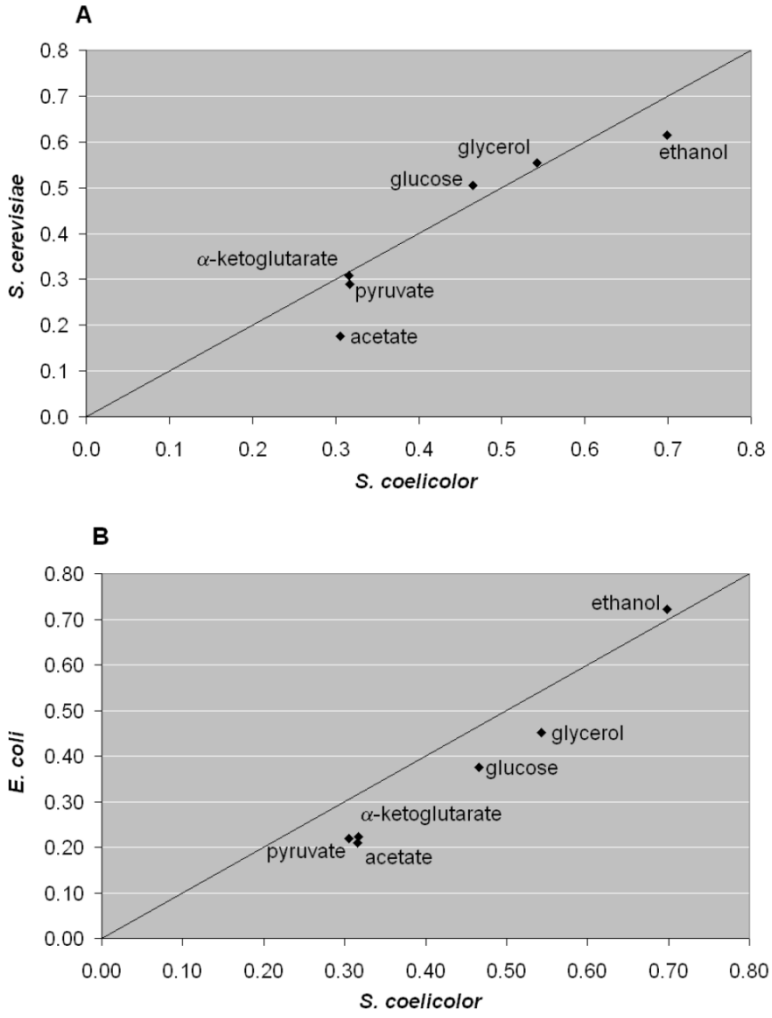
[1] The data are for *Streptomyces lividans*

3.3.7.2 Biomass yield

The biomass yields of *in silico S. coelicolor* A3(2) on various carbon sources were similar to that of *in silico* yeast (Förster et al. 2003a) (Figure 3-6A). The energy

requirements were considered for the simulations. The exception was growth on C-2 compounds like acetate and ethanol where *S. coelicolor A3(2)* was more efficient at utilizing these substrates. This is due to the action of the acetate kinase (encoded by *ackA*) and the phosphate acetyltransferase (encoded by *pta*), which catalyze the conversion of acetate into acetyl-phosphate and further into acetyl-CoA at the expense of only one high-energy phosphate bond hydrolysis in ATP. In eukaryotes these enzymes are not present and acetate is activated by acetyl-CoA synthase (encoded by *acs*) with the co-current hydrolysis of ATP to AMP, which is energetically more costly. Comparison of the growth capabilities of *in silico S. coelicolor* and *in silico E. coli* (Edwards et al. 2000b) showed that the *S. coelicolor* had higher biomass yields, which was primarily due to higher maintenance energy requirements in *E. coli* (Figure 3-6B).

Figure 3-6. Maximal simulated biomass yield of *S. coelicolor* A3(2) as compared to *S. cerevisiae* (Förster et al. 2003a) (A) and *E. coli* (Edwards et al. 2000b) (B) on different carbon sources (g/g substrate). All the calculations were made for substrate uptake rate of 6 C-mmol/g DW/h. The maintenance energy requirement was considered.



3.3.7.3 Anaerobic growth

The inability of *Streptomyces* to grow anaerobically has been a puzzle for quite some time (Hodgson 2000). Generally the microbes are capable of growing anaerobically, as they may obtain their Gibbs free energy by substrate-level phosphorylation resulting in fermentative metabolism. The presence of lactate dehydrogenase in the sequenced genome of *S. coelicolor* A3(2) indicates that the organism should be capable of having a fully operational fermentative metabolism with lactate as a key fermentation product. Indeed, excretion of lactate has been observed for *S. griseus* grown under microaerophilic conditions (Hockenhull et al. 1954). Two reasons were suggested to explain the anaerobic *Streptomyces* paradox: either the organisms are sensitive to the fermentation products or/and there are some essential reactions that require oxygen.

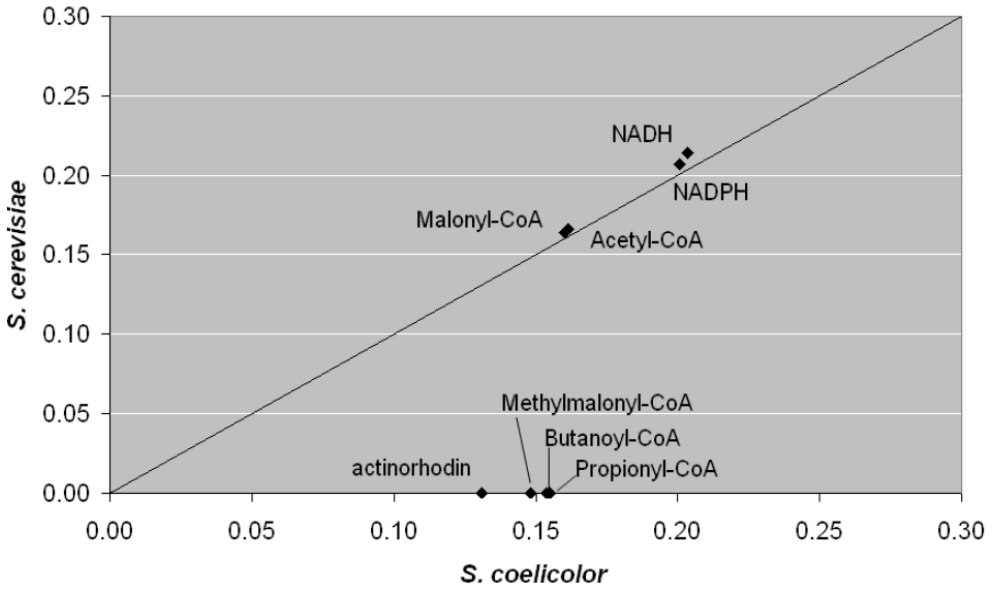
We analyzed the problem from the perspective of the reconstructed metabolic network and did not find any essential reactions that exclusively use oxygen. It seems that the essential dehydroorotate dehydrogenase E.C. 1.3.3.1 (SCO1482) required for pyrimidines biosynthesis and L-aspartate oxidase E.C. 1.4.3.16 (SCO3382) participating in nicotinamide nucleotides biosynthesis can both use oxygen and menaquinone as electron acceptors. Anaerobic growth can simply require that the produced menaquinol can be re-oxidized by the reverse action of succinate dehydrogenase. When glucose uptake rate was set to the maximal experimentally observed value (2.2 mmol/g DW/h) (Melzoch et al. 1997) and oxygen uptake rate was fixed to zero, the maximum specific growth rate was predicted to be 0.022 h^{-1} and 0.056 h^{-1} with ammonia and nitrate as nitrogen source respectively. The growth however implied conservation of a proton potential over the membrane, i.e. the protons exported during secretion of succinate are used for uptake of glucose. This is an unlikely situation *in vivo* and then substrate uptake might be the limiting factor for growth without oxygen. Glucose is transported into the *S. coelicolor* A3(2) by proton symport and the proton gradient is created primary in the respiratory process. Meanwhile, facultative anaerobes like yeast and *E. coli* use respectively facilitated diffusion and the PTS transport system, which remain active under anaerobic conditions.

3.3.7.4 Actinorhodin production

Several recent publications are dedicated to expressing bacterial and fungal polyketide synthase (PKS) genes in heterologous hosts. (Kealey et al. 1998; Kennedy et al. 2003; Kinoshita et al. 2003). As an example we addressed the issue of metabolic capabilities of yeast and *S. coelicolor* A3(2) to produce reducing cofactors and the most common polyketides precursors (Figure 3-7).

The capabilities of the metabolic networks of yeast and *S. coelicolor* A3(2) to produce acetyl-CoA, malonyl-CoA and reducing equivalents NADH and NADPH turned out to be very similar. However *Streptomyces* have a clear advantage in the form of more extensive polyols metabolism, which allows them to make such typical building blocks for polyketides production as methylmalonyl-CoA, propionyl-CoA and butanoyl-CoA. Development of heterologous production of certain polyketides from *E. coli* or yeast is tedious as it requires not only addition of PKS genes, but also the genes necessary for biosynthesis of precursors.

Figure 3-7. Predicted polyketides production capabilities of *S. coelicolor* A3(2) and *Saccharomyces cerevisiae* (Förster et al. 2003a) (in mol/mol glucose). All the calculations were made for glucose uptake rate of 1 mmol/g DW/h.



3.3.8 Filling the gaps in genome annotation

We assumed that the following processes had to be included in order to produce a "viable" model cell:

- transport reactions (nutrients and oxygen uptake, metabolites excretion);
- biosynthesis of the 12 biomass precursors;
- biosynthesis of each biomass component (amino acids, vitamins, co-factors, etc.);
- degradation of the experimentally utilizable substrates.

If any of these conditions were not fulfilled, a missing reaction (-s) was added according to the pathway structure (organism specific or general if the former was not available). 205 reactions without assigned ORFs were added to the model: 79 enzymatic reactions, 117 transport reactions, 5 spontaneous and 4 artificial reactions like maintenance, biomass assembly (Appendix 3.10). Out of the 79 enzymatic reactions, 27 were identified as being essential for growth on glucose and salts. In the context of functional genomics the reconstructed metabolic network hence serves as physiological evidence for the presence of these genes and allows directed search for the ORFs with the necessary function. Each of these 27 reactions was subsequently analyzed on an individual basis. The genome was searched for genes with specific protein motifs (http://www.sanger.ac.uk/Software/Pfam/) and the obtained hits were evaluated using the functions of the neighboring genes and their possible organization in an operon. The previously described Bayesian method (Green et al. 2004) has essentially the same logics, but it does not account for the protein motifs and uses homology search instead, besides, it requires model implementation in Pathway Tools (Karp et al. 2002).

3.3.8.1 Phospholipids biosynthesis

One of the enzymes missing in the phospholipids biosynthesis is cardiolipin synthase. No homologues of bacterial cardiopin synthase, which catalyzes the condensation of two phosphatidylglycerols into cardiolipin and glycerol were found in the genome of *S. coelicolor* A3(2). Presumably the cardiolipin synthesis occurs in the same way as in eukaryotes in reaction of phosphatidylglycerol with CDP-glycerol accompanied by formation of CMP. In this case the enzyme should possess a CDP-alcohol phosphatidyltransferase domain. With Pfam motif search 8 ORFs were found to contain such a domain in the *S. coelicolor* A3(2) genome. Three of these ORFs (SCO1527, SCO1389 and SCO5753) were synteny conserved with *M. tuberculosis* enzymes coding for phosphatidylinositol synthase (*pgsA*), cardiolipin synthase (*pgsA2*), and phosphatidylglycerol synthase (*pgsA3*) respectively (Jackson et al. 2000). Once a link has been established between gene and enzymatic function, it is possible to transfer the knowledge between different species by homology and synteny information, and we hereby predict that SCO5753 encodes phosphatididylglycerol synthase, SCO1527 – a phosphotidylinositol synthase and SCO1389 – a cardiolipin synthase.

3.3.8.2 Polyprenoids biosynthesis

The annotation of the polyprenoid biosynthesis was re-examined because 5 essential genes were found to be missing in the KEGG database, i.e. (E)-4-hydroxy-3-methylbut-2-enyl diphosphate reductase (IspH/LytB), pentaprenyl diphosphate synthase, hexaprenyl diphosphate synthase, octaprenyl diphosphate synthase, and nonaprenyl diphosphate synthase. The enzymes involved in the polyprenoid biosynthesis all have a prenyltransferase domain IPR001441. *S. coelicolor* A3(2) contains 7 genes with a polyprenyltransferase domain, i.e. SCO0185, SCO0565, SCO0568, SCO2509, SCO3858, SCO4583, SCO5250, and SCO6763. Two of the genes, SCO2509 and SCO3858, are synteny conserved with *M. tuberculosis* and must therefore have an important function. Furthermore SCO0185, SCO4583, SCO5250, and SCO6763 are synteny conserved with *S. avermitilis*. These 6 conserved genes are likely to be involved in general polyprenoid biosynthesis. The SCO0185 (*crtB*) is involved in biosynthesis of a secondary carotenoid metabolite (Lee et al. 2001), which is probably connected to the response to UV-radiation. This indicates that SCO0185 might not be true synteny conserved between *S. coelicolor* A3(2) and *S. avermitilis* but is a results of a transfer of a gene cluster. The SCO6763 is placed in another secondary metabolite gene cluster responsible for hopanoid biosynthesis, which is thought to be involved in the stress response in aerial mycelia by decreasing the water permeability of the plasma-membrane (Poralla et al. 2000). The last four unaccounted genes must be the genes that are responsible for polyprenoid biosynthesis for general purposes, i.e. plasma membrane electron carrier and transmembrane sugar carrier. The homologue of SCO3858 (Rv1086) has been biochemically characterized in *M. tuberculosis* as an omega,E,Z-farnesyl diphosphate synthase (Schulbach et al. 2000).

3.4 CONCLUSIONS

We have reconstructed the metabolic network of *S. coelicolor* A3(2). Besides resulting in improved annotation of several genes and suggestions for annotation of other genes, the reconstructed network may be used as a model of the metabolism in *Streptomyces.* Earlier attempts to model *Streptomyces* spp. (Roubos 2002; Bruheim

et al. 2002; Avignone et al. 2002) show that there is a demand for a high quality metabolic model of *Streptomyces* both in academic and industrial research. The model presented in this paper has been reconstructed based on the up-to-date knowledge about *Streptomyces* and can readily (or with small adjustments for species specificity) be used for simulations. We therefore expect that the model will be a useful tool for genome-wide analysis of the transcriptome and the proteome as well as for designing strategies for improvement of antibiotics yields in *Streptomyces*.

3.5 Methods

3.5.1 Reconstruction process

For reconstruction of *S. coelicolor* A3(2) metabolic network we used the annotated genome databases (KEGG PATHWAY database: http://www.genome.ad.jp/dbget-bin/get_htext?S.*coelicolor*.kegg+-f+T+w+C and The Wellcome Trust Sanger Institute database: http://www.sanger.ac.uk/Projects/S_*coelicolor*/scheme.shtml), metabolic databases (KEGG Ligand database http://www.genome.ad.jp/kegg/ligand.html, ExPASy Biochemical Pathways http://www.expasy.org/cgi-bin/search-biochem-index, ExPASy Enzyme Database http://www.expasy.org/enzyme, Swiss-Prot database http://www.expasy.org/sprot/sprot-top.html), biochemistry books (Ingraham et al. 1983; Stanier et al. 1986; Michal 1999), reviews (Hodgson 2000) and journal publications. The reconstruction process was started by downloading the *S. coelicolor* pathway database from KEGG ftp server (ftp://ftp.genome.ad.jp/pub/kegg/pathways/sco/) and the reactions corresponding to the given E.C. numbers/enzyme names were written using metabolic databases indicated above. Water, intracellular protons and hydroxyl ions were not included in the model, assuming that different non-enzymatic processes also use these substrates and therefore they don't need to be balanced in the set of enzymatic reactions. The functions of the included genes were compared to the functional assignments in The Wellcome Trust Sanger Institute database and when possible confirmed by information from the literature. In case the reaction was necessary to produce a viable *in silico* cell, but the corresponding gene was not in present in the annotated genome, the reaction was included in the model without genomic evidence. Unless the information of irreversibility of a reaction was available it was set as reversible. A list of the reactions with corresponding references is available as the supplementary material (Appendix 3.2). The biomass equation was made based on the macromolecular composition of *S. coelicolor* A3(2) (Shahab et al. 1996) and related species and is also available as the supplementary material (Appendix 3.5).

3.5.2 Modeling

Flux balance analysis (FBA) was used for quantification of metabolic fluxes (Varma et al. 1994; Schilling et al. 1999). The reactions set composed a stoichiometric matrix S with dimensions $m \times n$, where m was the number of metabolites and n was the number of reactions. Assuming pseudo-steady state over intracellular metabolites concentrations the metabolic model could be defined by the following matrix equation:

$$S \cdot v = 0, \qquad (2)$$

where v represented all the metabolic fluxes.

As the number of metabolic fluxes exceeded the number of mass balance constraints, there existed a set of feasible metabolic fluxes distributions. A solution

was found using linear programming by introducing an optimization problem: MAXIMIZE $Z = c \cdot v$, where c was a row vector showing the influence of individual fluxes on the objective function. Additional constraints were imposed on fluxes: $a_i \le v_i \le b_i$, indicating the reaction irreversibility ($0 \le v_i < \infty$) or measured flux. The transport fluxes for phosphate, sulfate, ammonia and oxygen were not restrained ($-\infty \le v_i < \infty$). The uptake rates of metabolites that were not available in the medium were set to zero ($v_i = 0$). The excretion of the major metabolic products (carbon dioxide, pyruvate, acetate, etc.) was allowed ($0 \le v_i < \infty$).

For simulations of growth the flux to biomass was set to a certain value and the substrate uptake rate was minimized. For simulation of antibiotic production, the substrate uptake rate was constrained and the flux towards the antibiotic was maximized.

In order to find fluxes that were active during alternative optimal solutions, the following algorithm was used:

- substrate uptake rates were constrained and the biomass growth rate was maximized;
- the growth rate was fixed to the found maximal value, thus an additional constrain was introduced;
- by changing vector c each reaction in the network was maximized and then minimized, if the flux through the reaction was different from zero in at least one of these optimizations, the reaction was considered to be active in alternative optimal solutions.

The calculations were performed using the commercially available linear programming package LINDO (Lindo Systems Inc., Chicago). The algorithm for finding fluxes that are active during alternative optimal solutions was implemented in Matlab (The MathWorks Inc., Natick), but LINDO package was used as the solver for the linear programming problems.

3.5.3 Estimation of energetic parameters

Knowing one of the energetic parameters (P/O ratio, Y_{xATP} or m_{ATP}) allows calculation of the other two parameters from continuous cultivations data. Because none of the parameters was known, they were estimated for P/O ratios in the range of 0.5 to 2 (Figure 3-2).

The Y_{xATP} is composed of three parts:

- ATP costs for the synthesis of biomass building blocks (amino acids, nucleotides, etc.), which was accounted for through reactions stoichiometry;
- ATP costs for polymerization of monomers into biological polymers (proteins, DNA, etc.), which was assumed to be the same as for *E. coli* (Ingraham et al. 1983) and was included in the reactions of macromolecules biosynthesis;
- and ATP costs for growth associated maintenance $Y_{xATP\_growth\_maintenance}$, added to the growth equation.

The last composite $Y_{xATP\_growth\_maintenance}$ as well as maintenance ATP (m_{ATP}) were not known and were estimated from the experimental chemostate data (Melzoch et al. 1997).

Normally the glucose uptake rate q_{gluc}, carbon dioxide production rate q_{CO_2} and oxygen uptake rate q_{O_2} are linearly dependent on the specific growth rate (Figure 3-3). The $Y_{xATP\_growth\_maintenance}$ and m_{ATP} were set up to arbitrary values and the simulations were run for each of the experimentally investigated dilution rate by fixing the specific growth rate and actinorhodin production rate to the experimental values and performing linear optimization for glucose uptake rate minimization (Appendix 3.6). The obtained q_{gluc}, q_{CO_2} and q_{O_2} dependence on dilution rate was compared to the experimental. The $Y_{xATP\_growth\_maintenance}$ and m_{ATP} were changed until a good prediction was obtained. For the further simulations the maximal P/O ratio was fixed to 1.5 and Y_{xATP} and m_{ATP} were set to the corresponding values.

3.6 Acknowledgements

We thank Prof. D.A. Hodgson for providing his Ph.D. thesis and for the excellent review on the primary metabolism of *Streptomyces*. We are grateful to Jochen Förster for sharing his experience in genome-scale modeling.

3.7 References

Avignone, R.C., White, J., Kuiper, A., Postma, P.W., Bibb, M., and Teixeira de Mattos, M.J. 2002. Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*. *Metab Eng* **4**:138-150

Barona-Gomez, F. and Hodgson, D.A. 2003. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* **4**:296-300

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., et al. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**:141-147

Blanco, J., Coque, J.J., and Martin, J.F. 1998. The folate branch of the methionine biosynthesis pathway in *Streptomyces lividans*: disruption of the 5,10-methylenetetrahydrofolate reductase gene leads to methionine auxotrophy. *J.Bacteriol.* **180**:1586-1591

Bott, M. and Niebisch, A. 2003. The respiratory chain of *Corynebacterium glutamicum*. *J.Biotechnol.* **104**:129-153

Bruheim, P., Butler, M., and Ellingsen, T.E. 2002. A theoretical analysis of the biosynthesis of actinorhodin in a hyper-producing *Streptomyces lividans* strain cultivated on various carbon sources. *Appl.Microbiol.Biotechnol.* **58**:735-742

Bucca, G., Brassington, A.M., Hotchkiss, G., Mersinias, V., and Smith, C.P. 2003. Negative feedback regulation of *dnaK*, *clpB* and *lon* expression by the DnaK chaperone machine in *Streptomyces coelicolor*, identified by transcriptome and *in vivo* DnaK-depletion analysis. *Mol.Microbiol.* **50**:153-166

Burgard, A.P., Pharkya, P., and Maranas, C.D. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol.Bioeng.* **84**:647-657

Cayrol, C., Petit, C., Raynaud, B., Capdevielle, J., Guillemot, J.C., and Defais, M. 1995. Recovery of respiration following the SOS response of Escherichia coli requires RecA-mediated induction of 2-keto-4-hydroxyglutarate aldolase. *Proc.Natl.Acad.Sci.U.S.A* **92**:11806-11809

Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., and Palsson, B.O. 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**:92-96

Duarte, N.C., Herrgard, M.J., and Palsson, B. 2004. Reconstruction and Validation of *Saccharomyces cerevisiae* iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Res.* **14**:1298-1309

Edwards, J.S., Ibarra, R.U., and Palsson, B.O. 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat.Biotechnol.* **19**:125-130

Edwards, J.S. and Palsson, B.O. 2000a. Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC.Bioinformatics.* **1**:1

Edwards, J.S. and Palsson, B.O. 2000b. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc.Natl.Acad.Sci.U.S.A* **97**:5528-5533

Famili, I., Förster, J., Nielsen, J., and Palsson, B.O. 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc.Natl.Acad.Sci.U.S.A* **100**:13134-13139

Fink, D., Falke, D., Wohlleben, W., and Engels, A. 1999. Nitrogen metabolism in *Streptomyces coelicolor* A3(2): modification of glutamine synthetase I by an adenylyltransferase. *Microbiology* **145**:2313-2322

Flett, F., Platt, J., and Cullum, J. 1987. DNA rearrangements associated with instability of an arginine gene in *Streptomyces coelicolor* A3(2). *J.Basic Microbiol.* **27**:3-10

Förster, J., Famili, I., Fu, P., Palsson, B.O., and Nielsen, J. 2003a. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**:244-253

Förster, J., Famili, I., Palsson, B.O., and Nielsen, J. 2003b. Large-scale evaluation of *in silico* gene deletions in *Saccharomyces cerevisiae*. *OMICS.* **7**:193-202

Garrity, G.M. 2002. *Bergey's Manual of Systematic Bacteriology*. Springer-Verlag, New York.

Green, M.L. and Karp, P.D. 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC.Bioinformatics.* **5**:76

Gupta, S.C. and Dekker, E.E. 1984. Malyl-CoA formation in the NAD-, CoASH-, and alpha-ketoglutarate dehydrogenase-dependent oxidation of 2-keto-4-hydroxyglutarate. Possible coupled role of this reaction with 2-keto-4-hydroxyglutarate aldolase activity in a pyruvate-catalyzed cyclic oxidation of glyoxylate. *J.Biol.Chem.* **259**:10012-10019

Hesketh, A.R., Chandra, G., Shaw, A.D., Rowland, J.J., Kell, D.B., Bibb, M.J., and Chater, K.F. 2002. Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*. *Mol.Microbiol.* **46**:917-932

Hockenhull, D.J., Fantes, K.H., Herbert, M., and Whitehead, B. 1954. Glucose utilization by *Streptomyces griseus*. *J.Gen.Microbiol.* **10**:353-370

Hodgson, D. A. Carbohydrate utilization in *Streptomyces coelicolor* A3(2). PhD thesis. 1980. University of East Anglia, Norwich, UK.

Hodgson, D.A. 2000. Primary metabolism and its control in *Streptomycetes*. *Advances in Microbial Physiology* **42**:47-238

Hopwood, D.A. 1999. Forty years of genetics with *Streptomyces*: from *in vivo* through *in vitro* to *in silico*. *Microbiology* **145**:2183-2202

Hostalek, Z., Tobek, I., Bobyk, M.A., and Kulayev, I.S. 1976. Role of ATP-glucokinase and polyphosphate glucokinase in *Streptomyces aureofaciens*. *Folia Microbiol.(Praha)* **21**:131-138

Hu, D.S., Hood, D.W., Heidstra, R., and Hodgson, D.A. 1999a. The expression of the *trpD*, *trpC* and *trpBA* genes of *Streptomyces coelicolor* A3(2) is regulated by growth rate and growth phase but not by feedback repression. *Mol.Microbiol.* **32**:869-880

Huang, G., Lih, C.-J., Pan, K.-H., and Cohen, S.N. 2001. Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes and Development* **15**:3183-3192

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**:929-934

Ihmels, J., Bergmann, S., and Barkai, N. 2004. Defining transcription modules using large-scale gene expression data. *Bioinformatics.* **20**:1993-2003

Inbar, L. and Lapidot, A. 1991. 13C nuclear magnetic resonance and gas chromatography-mass spectrometry studies of carbon metabolism in the actinomycin D producer *Streptomyces parvulus* by use of 13C-labeled precursors. *J.Bacteriol.* **173**:7790-7801

Ingraham, J.L., Maaløe, O., and Neidhardt, F.C. 1983. *Growth of the bacterial cell*. Sinauer Associates, Inc., Massachusetts.

Jackson, M., Crick, D.C., and Brennan, P.J. 2000. Phosphatidylinositol is an essential phospholipid of mycobacteria. *J.Biol.Chem.* **275**:30092-30099

Karp, P.D., Paley, S., and Romero, P. 2002. The Pathway Tools software. *Bioinformatics.* **18**:S225-S232

Kealey, J.T., Liu, L., Santi, D.V., Betlach, M.C., and Barr, P.J. 1998. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc.Natl.Acad.Sci.U.S.A* **95**:505-509

Kell, D.B. 2004. Metabolomics and systems biology: Making sense of the soup. *Curr.Opin.Microbiol.* **7**:296-307

Kennedy, J., Murli, S., and Kealey, J.T. 2003. 6-Deoxyerythronolide B analogue production in *Escherichia coli* through metabolic pathway engineering. *Biochemistry* **42**:14342-14348

Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. 2000. General introduction to actinomycete biology. In Practical *Streptomyces* Genetics , pp. 1-42. The John Innes Foundation, Norwich.

Kinoshita, K., Pfeifer, B.A., Khosla, C., and Cane, D.E. 2003. Precursor-directed polyketide biosynthesis in *Escherichia coli*. *Bioorg.Med.Chem.Lett.* **13**:3701-3704

Lee, H.S., Ohnishi, Y., and Horinouchi, S. 2001. A sigma B-like factor responsible for carotenoid biosynthesis in *Streptomyces griseus*. *J.Mol.Microbiol.Biotechnol.* **3**:95-101

Melzoch, K., Teixeira de Mattos, M.J., and Neijssel, O.M. 1997. Production of actinorphodin by *Streptomyces coelicolor* A3(2) grown in chemostat culture. *Biotechnology and Bioengineering* **54**:577-582

Michal, G. 1999. *Biochemical Pathways: An Atlas of Biocehmistry and Molecular Biology*. John Wiley & Sons, Inc., New York.

Navarrete, R.M., Vara, J.A., and Hutchinson, C.R. 1990. Purification of an inducible L-valine dehydrogenase of *Streptomyces coelicolor* A3(2). *J.Gen.Microbiol.* **136**:273-281

Novotna, J., Vohradsky, J., Berndt, P., Gramajo, H., Langen, H., Li, X.M., Minas, W., Orsaria, L., Roeder, D., and Thompson, C.J. 2003. Proteomic studies of diauxic lag in the differentiating prokaryote *Streptomyces coelicolor* reveal a regulatory network of stress-induced proteins and central metabolic enzymes. *Mol.Microbiol.* **48**:1289-1303

Pandya, K.P. and King, H.K. 1966. Ubiquinone and menaquinone in bacteria: a comparative study of some bacterial respiratory systems. *Arch.Biochem.Biophys.* **114**:154-157

Papp, B., Pal, C., and Hurst, L.D. 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**:661-664

Poralla, K., Muth, G., and Hartner, T. 2000. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol.Lett.* **189**:93-95

Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.O. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4**:R54

Roubos, H. Bioprocess modeling and optimization. Fed-batch clavulanic acid production by *Streptomyces clavuligerus*. PhD thesis. 2002. Ponsen & Looijen, Netherlands, Delft University of Technology.

Schilling, C.H., Covert, M.W., Famili, I., Church, G.M., Edwards, J.S., and Palsson, B.O. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J.Bacteriol.* **184**:4582-4593

Schilling, C.H., Edwards, J.S., and Palsson, B.O. 1999. Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol.Prog.* **15**:288-295

Schilling, C.H. and Palsson, B.O. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J.Theor.Biol.* **203**:249-283

Schulbach, M.C., Brennan, P.J., and Crick, D.C. 2000. Identification of a short (C15) chain Z-isoprenyl diphosphate synthase and a homologous long (C50) chain isoprenyl diphosphate synthase in Mycobacterium tuberculosis. *J.Biol.Chem.* **275**:22876-22881

Shahab, N., Flett, F., Oliver, S.G., and Butler, P.R. 1996. Growth rate control of protein and nucleic acid content in *Streptomyces coelicolor* A3(2) and *Escherichia coli* B/r. *Microbiology* **142**:1927-1935

Stanier, R.Y., Ingraham, J.L., Wheelis, M.L., and Painter, P.R. 1986. *General Microbiology*, Fifth edition. Macmillan Eduction LTD, UK.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E.D. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**:190-193

Stouthamer, A.H. and Bettenhaussen, C. 1973. Utilization of energy for growth and maintenance in continuous and batch cultures of microorganisms. A reevaluation of the method for the determination of ATP production by measuring molar growth yields. *Biochim.Biophys.Acta* **301**:53-70

Tang, L. and Hutchinson, C.R. 1993. Sequence, transcriptional, and functional analyses of the valine (branched-chain amino acid) dehydrogenase gene of *Streptomyces coelicolor*. *J.Bacteriol.* **175**:4176-4185

Thomas, M.G., Burkart, M.D., and Walsh, C.T. 2002. Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodigiosin and pyoluteorin biosynthesis. *Chem.Biol.* **9**:171-184

van Wezel, G.P., White, J., Bibb, M.J., and Postma, P.W. 1997. The *malEFG* gene cluster of *Streptomyces coelicolor* A3(2): characterization, disruption and transcriptional analysis. *Mol.Gen.Genet.* **254**:604-608

Varma, A. and Palsson, B.O. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl.Environ.Microbiol.* **60**:3724-3731

Yassin, A.F., Brzezinka, H., Schaal, K.P., Truper, H.G., and Pulverer, G. 1988. Menaquinone composition in the classification and identification of aerobic actinomycetes. *Zentralbl.Bakteriol.Mikrobiol.Hyg.* **267**:339-356

Zhang, Y.X., Tang, L., and Hutchinson, C.R. 1996. Cloning and characterization of a gene (*msdA*) encoding methylmalonic acid semialdehyde dehydrogenase from *Streptomyces coelicolor*. *J.Bacteriol.* **178**:490-495

3.8 Web references

http://www.genome.ad.jp/dbget-bin/get_htext?S.coelicolor.kegg+-f+T+w+C; *Streptomyces coelicolor* KEGG Pathways database
ftp://ftp.genome.ad.jp/pub/kegg/pathways/sco/; *Streptomyces coelicolor* KEGG pathway database (on ftp server)
http://www.genome.ad.jp/kegg/ligand.html; KEGG Ligand database
http://www.expasy.org/cgi-bin/search-biochem-index; ExPASy Biochemical Pathways database
http://www.sanger.ac.uk/Projects/S_*coelicolor*/scheme.shtml; The Wellcome Trust Sanger Institute database of *Streptomyces coelicolor*
http://www.expasy.org/enzyme; ExPASy Enzymes nomenclature database
http://www.expasy.org/sprot/sprot-top.html; Swiss-Prot protein knowledgebase
http://www.sanger.ac.uk/Software/Pfam/; Protein families database of alignments and HMMs
http://dbkweb.ch.umist.ac.uk/StreptoBASE/s_coeli/referencegel/; *Streptomyces coelicolor* 2D gel protein database

CHAPTER 4

Automated reconstruction of pathogenic *Streptomyces scabies* metabolism

Irina Borodina[1], Pichahpuk Uthaipaisanwong[2], Stephen Bentley[3], and Jens Nielsen[1]

[1] Center for Microbial Biotechnology, BioCentrum-DTU, Building 223, DK-2800 Kgs. Lyngby, Denmark
[2] King Mongkut's University of Technology Thonburi, Thailand
[3] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

4.1 Abstract

Common scab is a widespread disease of potatoes and other root and tubers crops, causing serious economical losses worldwide. Scab is caused by phytopathogenic *Streptomyces* species, of which *S. scabies* is the most common. Based on the recently released genome sequence of *S. scabies* we reconstructed a genome-scale metabolic network.
For making the network we used two approaches: one was a partly automated reconstruction using software package Pathway Tools, another was reconstruction by homology based on *S. coelicolor* A3(2) model [1].
Using Pathway Tools we could quickly generate a template model and predict protein complexes, transcriptional units, and transporters as well as identify 462 enzymes that are necessary for cellular function, but haven't been annotated (so called pathway holes). Using the prediction tools in-built in Pathway Tools we found the coding genes for 72 of the pathway holes.
By comparing the genomes of *S. scabies*, *S. coelicolor* and *S. avermitilis*, we reconstructed a stoichiometric model of *S. scabies*. In comparison to *S. coelicolor*, it lacked isocitrate lyase, but included thaxtomin biosynthesis pathway. We show that this model can be used to simulate for example which enzymes are essential for thaxtomin biosynthesis and for cells viability at different conditions.

Supplementary material: the Pathway Tool model of S. scabies is supplied in Appendix 4.1 and the bioopt format of the model is given in Appendix 4.2.

Keywords

Streptomyces scabies, metabolic model, *in silico*

4.2 Introduction

Common scab of potatoes is caused by several *Streptomyces* species [2, 3] and occurs in most of potato cultivation areas. It can also appear on other tubers as beets, turnip, radish, carrots, etc. Scab is characterized by formation of lesions widely ranging in size, color and texture. Appearance of lesions is a large economical problem, because the market value of contaminated tubers decreases significantly. The main organism causing potato scab is *Streptomyces scabies*, which has recently been sequenced in collaboration between Welcome Trust Sanger Institute (UK) and Cornell University (USA). The sequence was made publicly available (ftp://ftp.sanger.ac.uk/pub/pathogens/ssc/) while the annotation is currently ongoing.
The availability of genomic sequence allows us to better understand pathogenicity mechanism of *S. scabies* and to identify the targets for fighting this pathogen. To achieve this aim, functional genomics, connecting the sequences of the genes to their function, is of the utmost importance. Metabolic models have previously proven to be useful in complementing the manual genome annotation process [4]. They allow fast identification of missing enzymes to ensure the complete functionality of the organism [1]; they also allow predictions of lethal or non-pathogenic phenotypes, helping to find cure against pathogens [5].
Here we report reconstruction of *S. scabies* metabolic model and its application for studying organism physiology.

4.3 Results and discussion

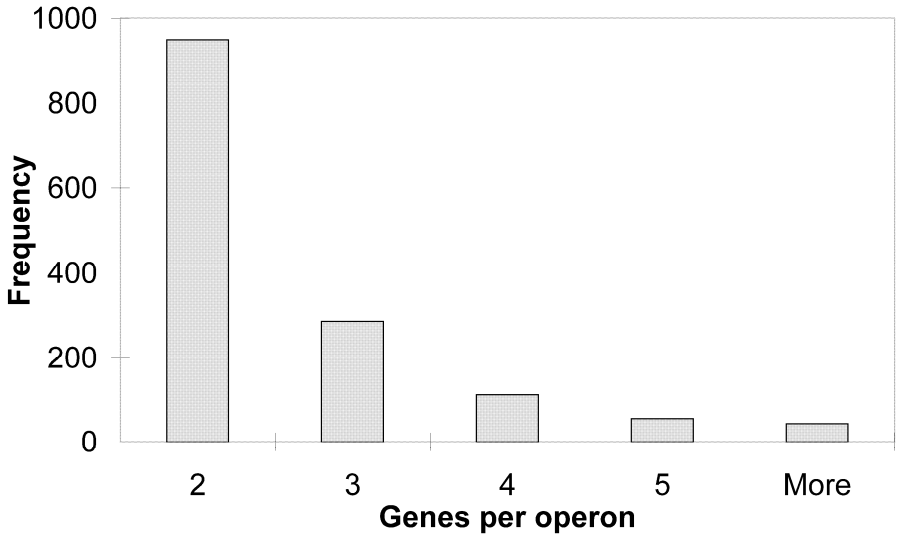
4.3.1 Automated reconstruction using genome annotation data

The complete genome sequence of 10,148,695 bp was initially automatically annotated using Glimmer 3.02 [6] (http://www.cbcb.umd.edu/software/glimmer). 9178 open reading frames were predicted, of which 9,106 encode proteins and 72 – transport RNAs. Of all the protein-coding ORFs, 1,759 were annotated as hypothetical proteins and further 1,971 did not match with any other proteins sequence in the database, meaning that 5,376 genes had some function annotated.
The output file from Glimmer 3.02 was converted into gene bank format file and used as an input into Pathway Tools version 10.5 (SRI International, USA). Pathway Tools uses an advanced algorithm to assign reactions to the genes, based on E.C. numbers and gene names from the annotation. The initial model, consisting of a thousand enzymatic reactions and 714 metabolites, was quite large considering that E.C. numbers were available only for 18 genes and most of the functions were found by unambiguous enzyme name matching (Table 4-1). Furthermore 1,297 proteins were predicted to be metabolic enzymes, but exact function could not be assigned to them because of ambiguity problem, non-standard enzyme names or other issues. It is possible to go through the list of these enzymes and assign their function manually later.
After creating the initial database we predicted protein complexes using a feature from Pathway Tools. The proteins that have the same function, have subunit indication in their name and are positioned adjacent to each other on the chromosome are grouped into protein complexes. We identified 43 protein complexes, which included NAD(P) transhydrogenase, urease, DNA gyrase, ATP synthase, acetolactate synthase and others.
Our next step was to predict transcriptional units. The program works by first finding adjacent genes that are transcribed in the same direction (directon). If a single gene is surrounded by genes that are transcribed in a different direction, it is automatically assigned as a transcriptional unit, because it would need its own promoter. In case of a longer directon, the genes in it are analyzed for their function, for belonging to the same pathway, for being parts of a protein complex, for base pair distances between them and then if the estimated score is higher than a threshold, they are assigned to be part of an operon. In our case the program predicted 6,798 transcriptional units, of which 1,446 were operons and the rest 5,352 were single genes. The lengths of the operons are shown on a histogram (Figure 4-1). The longer the operons, the less abundant they are in the genome.

Table 4-1. Statistics of the *S. scabies* pathway genome database at different steps of reconstruction in Pathway Tools.

	Initial	Create protein complexes	Transcription Units	Transport
Pathways	226	226	226	230
Enzymatic reactions	1000	1005	1005	1026
Transport reactions	2	2	3	**63**
Polypeptides	9113	9113	9113	9113
Protein complexes	1	**44**	44	50
Enzymes	678	711	711	784
Transporters	10	10	10	149
Compounds	714	721	721	747
Transcription units	0	0	**6798**	6796

Figure 4-1. Histogram shows frequency of operons with a specific number of genes in them.



A special program module - transporter identification parser - was used to predict transport proteins and their substrates, 146 high confidence transporters were predicted (Table 4-1).

As the final step, we searched if all metabolic pathways were complete using "pathway hole filler" [7]. From the genes that already have been included in the model, the estimation is made whether a certain pathway is present in the organism. For example, the presence of an enzyme which is unique for a pathway gives a higher score to this pathway existence. The enzymes, which are necessary for the pathway to work, but absent from the model are found. In case of *S. scabies* 462 pathway holes were detected (Appendix 4.3). The sequences for the corresponding enzymes were found in the genomes of *S. coelicolor* and *S. avermitilis* and were BLASTed against *S. scabies* genome to identify the candidates for filling in the holes. 72 candidate enzymes with reliability value of at least 0.9 were found (Appendix 4.4). Among these enzymes were GDP diphosphokinase (E.C. 2.7.6.5) and ppGpp diphosphatase (E.C. 3.1.7.2), which are responsible for maintaining the suitable levels of global regulator metabolite ppGpp. Both enzymatic functions were assigned to a single open reading frame SCAB07497, which is also consistent with observations in other Streptomycetes, where these activities are combined in one enzyme [8].

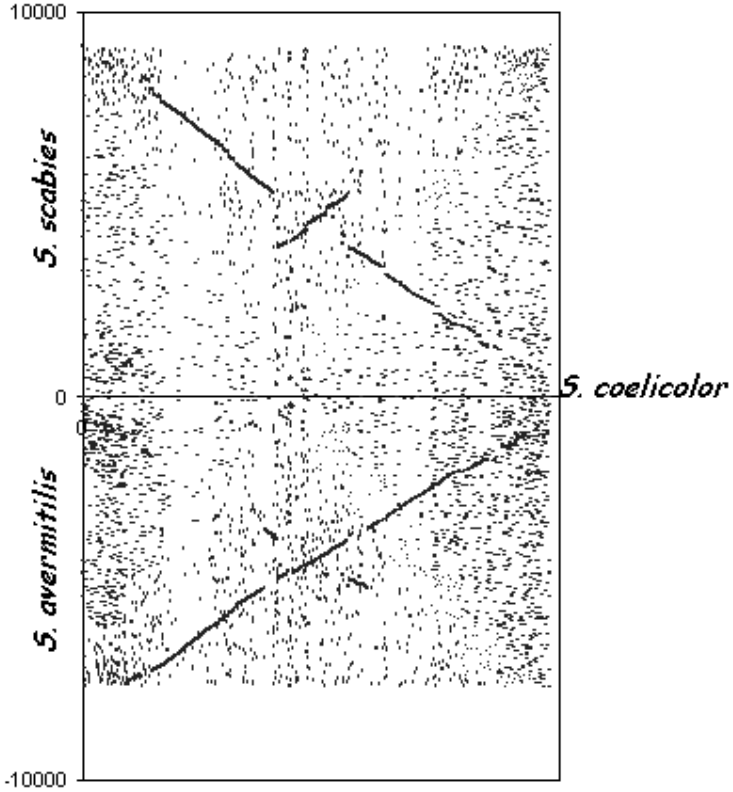
We exported the reactions from the database and formatted the file so it could be used as an input for stoichiometric modeling. However, the problems associated with automatic database reconstruction (see chapter 3) were numerous and multiple manual corrections needed to be done before the model could be used for simulations. We evaluated it as being inexpedient to make all the corrections at this stage, where we only had an automatic draft annotation of the genome to start with.

4.3.2 Reconstruction using homologous search to S. coelicolor genes

Because of the problems encountered above we decided to use a shortcut for construction of a stoichiometric model, namely by using the previously described *S. coelicolor* model as a template. Each gene of *S. coelicolor* [1] was BLASTed against

S. scabies open reading frames. The same procedure was repeated using *S. avermitilis* genes. There are a large number of genes, which function is conserved across all three *Streptomyces* species. For some of these genes also the position on the chromosomes is conserved, these synteny conserved genes lie on the diagonals of the plots (Figure 4-2).

Figure 4-2. Global comparison of genes in *S. coelicolor*, *S. avermitilis* and *S. scabies*. The dots mark the position of homologous genes on the chromosome of different species.



625 genes from the *S. coelicolor* model [1] were homologous across all three species. For 62 genes either no homologue was found in *S. scabies* when BLASTing with *S. avermitilis* genes (BLASTing with *S. coelicolor* found a homologue) or the closest homologue was different from that found by BLASTing with *S. coelicolor* genes. For 6 more genes homologues were only found when searching with *S. avermitilis* sequences. Eventually there were 18 *S. coelicolor* genes, which did not have a homologue in the *S. scabies* genome (Figure 4-3, Table 4-2).
Some of these 18 genes performed the same function as other genes that were found in *S. scabies*. The rest of the genes, which function could not be compensated, are shown in Table 4-2. These reactions were excluded from the *S. coelicolor* model to provide a draft model of *S. scabies*. Among these reactions were secondary metabolic reactions leading to actinorhodin, prodigiosins and calcium-dependent antibiotics. The secondary metabolites clusters often are species-specific, therefore it is not surprising they are not present in *S. scabies*, which is not known to produce pigmented antibiotics.

Figure 4-3. Statistics over metabolic genes homology across the three species. SCO – *S. coelicolor*, SAV – *S. avermitilis*, SCAB – *S. scabies*.

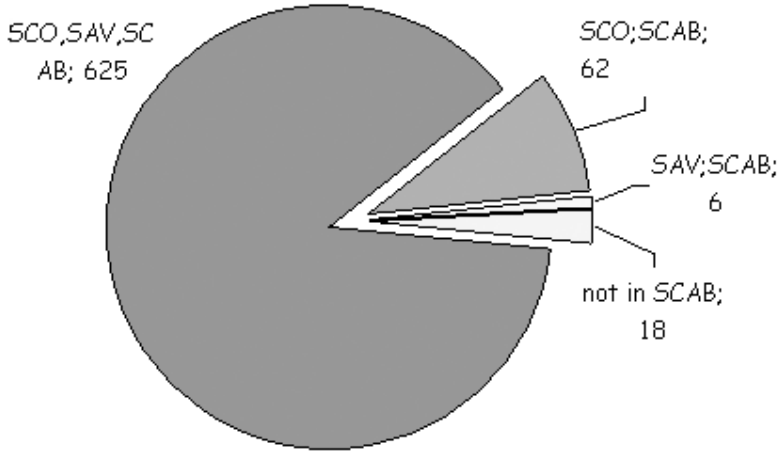


Table 4-2. Reactions from the *S. coelicolor* model, which were NOT included in the *S. scabies* model due to the absence of close homologous enzyme, which could catalyze the reaction.

***S. coelicolor* A3(2) ORFs without homologues in S. scabies genome**	**Function**	**E.C.#**	**Reactions**	**Essentiality under 63 conditions [1]**
SCO5078	actinorhodin biosynthesis		C16H16O5 + 0.5 O2 + NAD -> C16H14O6_2 + NADH	0
SCO5081	actinorhodin biosynthesis		C16H16O5 + 0.5 O2 + NAD -> C16H14O6_2 + NADH; C32H26O12 + O2 -> ACT	0
SCO5077 (SAV2084)	actinorhodin biosynthesis		C16H16O5 + 0.5 O2 + NAD -> C16H14O6_2 + NADH	0
SCO5079 (SAV0140)	actinorhodin biosynthesis		C16H16O5 + 0.5 O2 + NAD -> C16H14O6_2 + NADH	0
SCO5089 (SAV2373)	actinorhodin polyketide synthase acyl-carrier protein	2.3.1.-	8 MALCOA -> C16H17O8 + 8 COA + 8 CO2	0
SCO5090 (SAV2383)	actinorhodin polyketide synthase bifunctional cyclase/dehydratase		C16H20O9 -> C16H16O7	0
SCO5889	PCP - peptidyl carrier protein		PRO-AMP -> PRO-PCP + AMP	0
SCO7311	lysine decarboxylase	4.1.1.18	LYS -> CADA + CO2	2
SCO3215	Glutamate-3-methyltransferase		GLU + MTHF -> THF + NRP1	0
SCO0982 (SAV2043)	isocitrate lyase	4.1.3.1	ICIT -> GLX + SUCC	6
SCO2144 (SAV6487)	D-glycerate transport via sodium symport		DGLYCERATExt + Naxt -> DGLYCERATE + Na	1
SCO3029 (SAV5047)	L-lactate transport via proton symport/putative L-lactate permease		LLACxt + H <-> LLAC	1
SCO0985 (SAV2046)	5-methyltetrahydropteroyltriglutamate-homocysteine *S*-methyltransferase	2.1.1.14	HCYS + MTHPTGLU <-> THPTGLU + MET	0
SCO6075 (SAV1247)	3'-5'-cyclic nucleotide phosphodiesterase	3.1.4.17	cAMP -> AMP; cGMP -> GMP	0

Surprising was the absence of isocitrate lyase and cyclic nucleotide phosphodiesterase. Possibly these enzymes have quite different sequences in *S. coelicolor* and *S. avermitilis*. Isocitrate lyase, which lyses isocitrate into glyoxylate and malate, is the principal enzyme in glyoxylate shunt, a pathway that is responsible for generation of C4 metabolites from C2 compounds during growth on acetate, fatty acids or similar. There are, however, several other organisms that can grow on acetate in spite of the lack of isocitrate lyase activity, among them *Rhodobacter* spp. [9, 10], α-proteobacteria and *S. collinus* [11]. While several alternative routes from acetate-CoA to glyoxylate have been proposed, they haven't been explicitly proven. Cyclic nucleotide phosphodiesterase hydrolyses cyclic AMP and GMP into their normal form and hence regulates their level in the cell. cAMP and cGMP are important signaling molecules, they modulate DNA-binding of a transcription factor cAMP receptor protein [12]. Both adenylate cyclase, which is responsible for biosynthesis of cyclic nucleotides, and cAMP receptor proteins have homologues in *S. scabies*, SCAB03456 and SCAB04132 respectively. Deletion of either of these genes in *S. coelicolor* causes defects in germination and growth delay and affects sporulation [13]. As it is very unlikely that cyclic nucleotide phosphodiesterase is missing, we searched the *S. scabies* annotation and indeed we found a calcineurin-like phosphoesterase (SCAB02697), which is most likely to perform this function. The *S. coelicolor* phosphodiesterase is also calcineurin-like according to its Pfam domain (http://pir.georgetown.edu/pirwww/index.shtml).

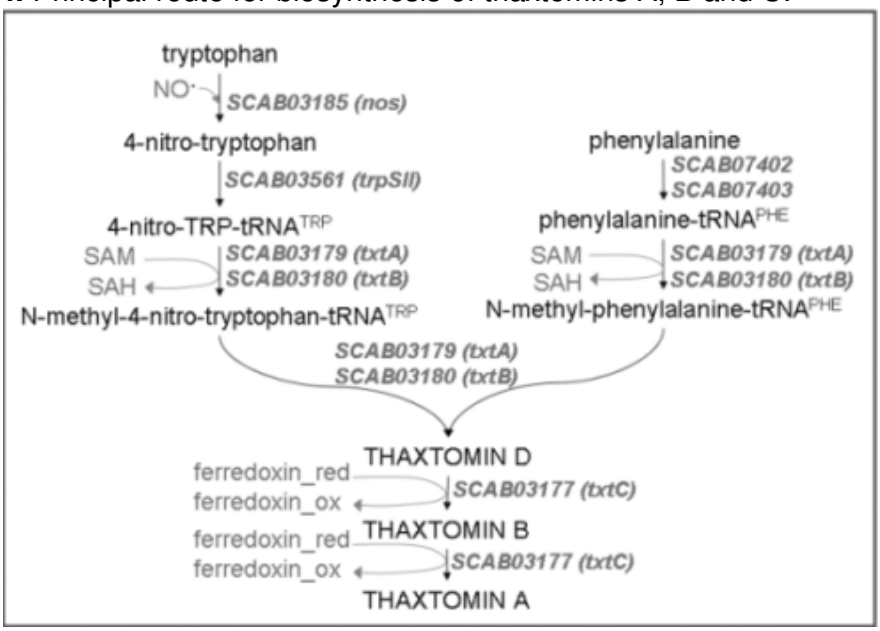
4.3.3 Reconstruction of thaxtomins biosynthesis

The pathogenicity of scab-causing *Streptomyces* species has so far been mainly attributed to production of phytotoxins thaxtomins. All *Streptomyces* species that are capable of causing scab can make thaxtomins [3]. Deletion of thaxtomins biosynthetic genes from pathogenic *Streptomyces* makes them incapable of colonizing tubers [14]. Thaxtomins are dipeptides, consisting of phenylalanine and nitrated tryptophan [15, 16] (Figure 4-4). Thaxtomins inhibit cellulose biosynthesis in expanding plant tissues, stimulate calcium ion spiking and cause cellular death [3]. Arabidopsis mutant resistant to thaxtomin was isolated; it had a mutation in *txr1* gene, which apparently is a regulator of thaxtomin transport. The resistant plant had reduced uptake rate of thaxtomin. [17]

Another *Streptomyces* gene *nec1* is also connected to pathogenicity. It was initially discovered by expression of *S. scabies* cosmid library in non-pathogenic *S. lividans* species [18]. The clones carrying *nec1* gene became capable of causing necrosis of potato slices. The gene is conserved between several parasitic *Streptomyces* species [19, 20]. *Nec1* gene encodes a secreted protein, however, the exact mechanism of this protein action is so far unknown [3].

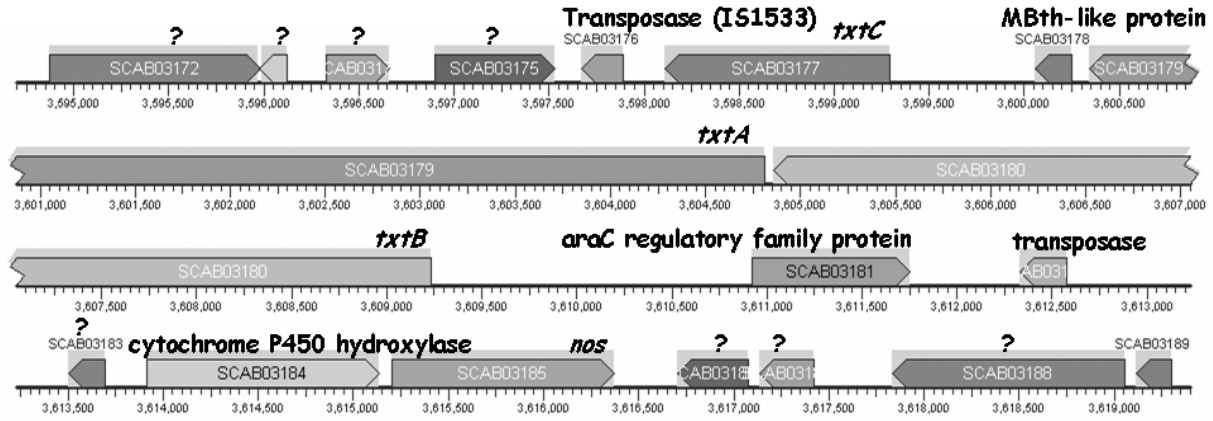
Thaxtomin genes and necrosis factor gene are located on a mobilizable pathogenicity island (PAI), which also often includes other supposedly virulence genes as cytokinin biosynthetic genes and saponinase [3]. The island transfers easily between *Streptomyces* species by conjugation, but it's not always that adoption of this island confers pathogenicity phenotype [21].

Figure 4-4. Principal route for biosynthesis of thaxtomins A, B and C.



The pathway is not present in MetaCyc database and hence was not included in the *S. scabies* model during automatic reconstruction. It was not annotated during the automatic genome annotation either. We used similarity search to the pathogenicity island of *S. turgidiscabies* (AY707079–AY707081) [21] to find thaxtomin biosynthetic genes in *S. scabies*. At least four genes are known to be directly involved in thaxtomin biosynthesis *txtA*, *txtB*, *txtC* and *nos*. Homologues of each of these genes were found in *S. scabies* with zero E-values (Figure 4-5). The sequence alignments of each gene are shown in Appendix 4.5.

Figure 4-5. Position of the putative thaxtomins biosynthetic genes on the chromosome of *S. scabies*.



The genes *txtC*, *txtA* and *txtB* are located adjacent to each other and are surrounded by transposases from up- and downstream. Two other genes are present in the same transposable element – a putative regulatory protein of AraC family and an MBth-like protein which was originally found in *Mycobacterium tuberculosis*, where it functions as salicylate synthase and is involved in mycobactin biosynthesis [22, 23], and later

was also found in some secondary metabolic genes clusters in Actinomycetes [24]. *araC* gene is positioned in opposite direction than the other genes and is most likely transcribed from a separate promoter. Nitric oxide synthase gene *nos* is positioned close to the transposon element, next to a putative cytochrome P450 hydroxylase, an enzyme likely involved in the secondary metabolism.

All the thaxtomins that have been isolated so far from infected plant tissues have 4-nitrotryptophyl in their structures. The nitration of tryptophan has been shown to occur in many *Streptomyces* species by the action of nitric oxide synthase (NOS). [25]. Bacterial nitric oxide synthase NOS performs formation of nitric oxide from arginine. Nitric oxide reacts with with tryptophan, generating 4-nitro-tryptophan. The positional targeting of nitration to the 4^{th} position is achieved by complex formation between NOS and an unusual tryptoptophanyl-tRNA synthetase (TrpRS II), at least in *Deinococcus radiodurans* [26-29]. The *trpSII* gene is present in both *S. coelicolor* (SCO4839) and *S. avermitilis* (SAV3417) genomes, and we also detected a homologue in *S. scabies* (SCAB03561). TrpRS II charges $tRNA^{Trp}$ with 4-nitro-tryptophan. Another amino acid, which composes the thaxtomin dipeptide, is tryptophan. It is also attached to a specific tRNA. Both amino acid precursors are methylated, most probably by S-adenosylmethionine (SAM), because *txtA* and *txtB* carry SAM-dependent N-methyltransferase domains [14]. The methylated precursors are joined by two peptide bonds into thaxtomin D by TxtA and TxtB. The P450 cytochrome TxtC hydroxylates thaxtomin D into thaxtomin B and further into thaxtomin A [30], most probably using ferredoxin as electron donor. The reactions that lead to thaxtomins biosynthesis are highlighted in Table 4-3. All these reactions were included in the model.

4.3.3 Simulations

Simulations were performed using linear programming software package LINDO API (LINDO Systems Inc., USA). An in-house program Bioopt was used as an interface to communicate the model stoichiometric matrix, flux limits and the optimization criteria to LINDO.

We simulated single gene deletions and searched for the ones that eliminate thaxtomin A biosynthesis. 38 reactions were found for growth on minimal medium (data not shown). However, in real situation the parasite might be capable of getting amino acids from the plant host, therefore we then simulated gene deletions during growth on glucose and all twenty amino acids. Besides the reactions from thaxtomin pathway, only 6 other reactions were found that can eliminate thaxtomin biosynthesis. Five of these reactions are involved in S-adenosinemethionine metabolism and are catalyzed by methionine adenosyltransferase, adenosylhomocysteinase, methionine synthase, 5,10-methylenetetrahydrofolate reductase and flavin reductase. The last reaction, catalyzed by ribose-phosphate diphosphokinase was involved in nucleotides biosynthesis. Unfortunately, all of these enzymes also have analogues in plants (we looked at *Arabidopsis thaliana*) and hence are not suitable targets for management of *S. scabies* infections.

Table 4-3. Enzymes necessary for thaxtomins biosynthesis in *S. scabies* according to computer simulations. The highlighted enzymes are specific for thaxtomin pathway.

Enzyme	Gene	Reaction	Medium with amino acids	
			Thaxtomins	Growth
tryptophanyl-tRNA synthase	SCAB03561 (*trpSII*)	4-nitro-TRP + ATP + tRNA_trp -> 4-nitro-TRP-tRNA_trp + PPI + AMP	0%	100%
nitric oxide synthetase	SCAB03185 (*nos*)	ARG + 2 O2 + 1.5 NADPH -> CITR + NO + 1.5 NADP	0%	100%
nitric oxide synthetase	SCAB03185 (*nos*)	NO + O2 + 0.5 NADH + TRP -> 4-nitro-TRP + 0.5 NAD	0%	100%
phenylalanyl-tRNA synthase	SCAB07402 (alpha subunit) AND SCAB07403 (beta subunit)	PHE + ATP + tRNA_phe -> PHE-tRNA_phe + PPI + AMP	0%	100%
SAM-dependent methyltransferase	SCAB03179 (*txtA*) AND SCAB03180 (*txtB*)	PHE-tRNA + SAM -> N-methyl-PHE-tRNA + SAH	0%	100%
SAM-dependent methyltransferase	SCAB03179 (*txtA*) AND SCAB03180 (*txtB*)	4-nitro-TRP-tRNA + SAM -> N-methyl-4-nitro-TRP-tRNA + SAH	0%	100%
non-ribosomal protein synthesis	SCAB03179 (*txtA*) AND SCAB03180 (*txtB*)	N-methyl-PHE-tRNA + N-methyl-4-nitro-TRP-tRNA -> THAXTOMIN_D + 2 tRNA	0%	100%
P450 cytochrome hydroxylase	SCAB03177 (*txtC*)	THAXTOMIN_D + FERO + O2 -> FERI + THAXTOMIN_B	0%	100%
P450 cytochrome hydroxylase	SCAB03177 (*txtC*)	THAXTOMIN_B + FERO + O2 -> FERI + THAXTOMIN_A	0%	100%
flavin reductase	SCAB08267 OR SCAB05029	FAD + NADH <-> FADH2 + NAD	0%	0%
methionine adenosyltransferase	SCAB07532	MET + ATP -> PPI + PI + SAM	0%	0%
ribose-phosphate diphosphokinase	SCAB05417 OR SCAB00918	R5P + ATP <-> PRPP + AMP	0%	0%
adenosylhomocysteinase	SCAB05519	SAH -> HCYS + ADN	0%	0%
methionine synthase	SCAB07333	HCYS + MTHF <-> THF + MET	1%	69%
5,10-methylenetetrahydrofolate reductase (FADH2)	SCAB06780	METTHF + FADH2 -> MTHF + FAD	1%	69%

4.4 Conclusions

In this work we performed fast reconstruction of the metabolic network of a plant parasite – *Streptomyces scabies*. We used two methods: automatic reconstruction of a template model using Pathway Tools and manual adaptation of an already available *S. coelicolor* A3(2) metabolic network [1].

Pathway Tools is useful for fast automated reconstruction of the sequenced organisms. It uses the genome annotation to assign reactions to genes, naturally, the higher the quality of the annotation, the better the model. The software has many nice features, as predictors of proteins complexes, transporters and transcriptional units. One of the most useful features is pathway hole finder, which allows identification of enzymes that have been missed during annotation for one or another reason. The software is useful for storing various information about enzymes in an easily accessible form. It allows quick publication of the model on the World Wide Web and has navigation options that would allow the users to browse the network. The pathway genome database reconstructed in Pathway Tools can be readily used for interpretation of –omic data. There's, however, one disadvantage: in our hands the construction of stoichiometric model from the Pathway Tools database was very tedious and time consuming. More features should be added to the program if it should be suitable for mathematical simulations.

Reconstruction of basic *S. scabies* model from the previously made *S. coelicolor* model was fast. A great advantage was that the reconstruction was based on homology and hence did not depend on the genome annotation. Most of the genes in *S. coelicolor* model had homologues in *S. scabies* and hence most of reactions were common for these two species. We had to remove a few reactions (actinorhodin and calcium-dependent antibiotics biosynthesis and isocitrate lyase) and we added biosynthesis of thaxtomins. The model was used to simulate which reactions were essential for thaxtomin production in environment, which contains amino acids. S-adenosinemethionine metabolism was found essential for thaxtomin biosynthesis.

In conclusion, the choice of the metabolic model reconstruction technique depends on the purpose the model will have to fulfill. For data storage and distribution, and -omic data interpretation reconstruction in Pathway Tools is a good option. For mathematical simulations, building up on the available models of related organisms can be more time efficient.

4.5 Acknowledgements

We thank Stephan Bentley for the idea of the project and for supplying us with draft genome annotation data. We also thank Rosemary Loria (Cornell University, USA) for discussion on the project. IB thanks Technical University of Denmark for PhD. scholarship.

4.6 References

1. **Borodina I, Krabben P, Nielsen J**. (2005). "*Genome-scale analysis of Streptomyces coelicolor* A3(2) metabolism." *Genome Res*, **15**(6):820-829.
2. **Takeuchi T, Sawada H, Tanaka F, Matsuda I**. (1996). "*Phylogenetic analysis of Streptomyces spp. causing potato scab based on 16S rRNA sequences*." *Int J Syst Bacteriol*, **46**(2):476-479.
3. **Loria R, Kers J, Joshi M**. (2006). "*Evolution of plant pathogenicity in Streptomyces*." *Annu Rev Phytopathol*, **44**:469-487.

4. **Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BO *et al***. (2006). "*Systems approach to refining genome annotation*." *Proc Natl Acad Sci U S A*, **103**(46):17480-17484.
5. **Raman K, Rajagopalan P, Chandra N**. (2005). "*Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs*." *PLoS Comput Biol*, **1**(5)
6. **Takano E, Gramajo HC, Strauch E, Andres N, White J, Bibb MJ**. (1992). "*Transcriptional regulation of the redD transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in Streptomyces coelicolor A3(2)*." *Mol Microbiol*, **6**(19):2797-2804.
7. **Green ML, Karp PD**. (2004). "*A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases*." *BMC Bioinformatics*, **5**:76
8. **Martinez-Costa OH, Fernandez-Moreno MA, Malpartida F**. (1998). "*The relA/spoT-homologous gene in Streptomyces coelicolor encodes both ribosome-dependent (p)ppGpp-synthesizing and -degrading activities*." *J Bacteriol*, **180**(16):4123-4132.
9. **Meister M, Saum S, Alber BE, Fuchs G**. (2005). "*L-malyl-coenzyme A/beta-methylmalyl-coenzyme A lyase is involved in acetate assimilation of the isocitrate lyase-negative bacterium Rhodobacter capsulatus*." *J Bacteriol*, **187**(4):1415-1425.
10. **Kornberg HL, Lascelles J**. (1960). "*The formation of isocitratase by the Athiorhodaceae*." *J Gen Microbiol*, **23**:511-517.
11. **Han L, Reynolds KA**. (1997). "*A novel alternate anaplerotic pathway to the glyoxylate cycle in streptomycetes*." *J Bacteriol*, **179**(16):5157-5164.
12. **Derouaux A, Dehareng D, Lecocq E, Halici S, Nothaft H, Giannotta F, Moutzourelis G, Dusart J, Devreese B, Titgemeyer F *et al***. (2004). "*Crp of Streptomyces coelicolor is the third transcription factor of the large CRP-FNR superfamily able to bind cAMP*." *Biochem Biophys Res Commun*, **325**(3):983-990.
13. **Derouaux A, Halici S, Nothaft H, Neutelings T, Moutzourelis G, Dusart J, Titgemeyer F, Rigali S**. (2004). "*Deletion of a cyclic AMP receptor protein homologue diminishes germination and affects morphological development of Streptomyces coelicolor*." *J Bacteriol*, **186**(6):1893-1897.
14. **Healy FG, Wach M, Krasnoff SB, Gibson DM, Loria R**. (2000). "*The txtAB genes of the plant pathogen Streptomyces acidiscabies encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity*." *Mol Microbiol*, **38**(4):794-804.
15. **King RR, Lawrence CH, Calhoun L**. (1989). "*Isolation and characterization of phytotoxins associated with Streptomyces scabies*." *J Chem Soc Chem Commun*, **13**:849-850.
16. **Lawrence CH, Clark M, King RR**. (1990). "*Induction of common scab symptoms in aseptically cultures potato tubers by the vivotoxin, thaxtomin*." *Phytopathology*, **80**:606-608.
17. **Scheible WR, Fry B, Kochevenko A, Schindelasch D, Zimmerli L, Somerville S, Loria R, Somerville CR**. (2003). "*An Arabidopsis mutant resistant to thaxtomin A, a cellulose synthesis inhibitor from Streptomyces species*." *Plant Cell*, **15**(8):1781-1794.

18. **Bukhalid RA, Loria R**. (1997). "*Cloning and expression of a gene from Streptomyces scabies encoding a putative pathogenicity factor*." *J Bacteriol*, **179**(24):7776-7783.
19. **Bukhalid RA, Takeuchi T, Labeda D, Loria R**. (2002). "*Horizontal transfer of the plant virulence gene, nec1, and flanking sequences among genetically distinct Streptomyces strains in the Diastatochromogenes cluster*." *Appl Environ Microbiol*, **68**(2):738-744.
20. **Bukhalid RA, Chung SY, Loria R**. (1998). "*nec1, a gene conferring a necrogenic phenotype, is conserved in plant-pathogenic Streptomyces spp. and linked to a transposase pseudogene*." *Mol Plant Microbe Interact*, **11**(10):960-967.
21. **Kers JA, Cameron KD, Joshi MV, Bukhalid RA, Morello JE, Wach MJ, Gibson DM, Loria R**. (2005). "*A large, mobile pathogenicity island confers plant pathogenicity on Streptomyces species*." *Mol Microbiol*, **55**(4):1025-1033.
22. **Zwahlen J, Kolappan S, Zhou R, Kisker C, Tonge PJ**. (2007). "*Structure and mechanism of MbtI, the salicylate synthase from Mycobacterium tuberculosis*." *Biochemistry*, **46**(4):954-964.
23. **Harrison AJ, Yu M, Gardenborg T, Middleditch M, Ramsay RJ, Baker EN, Lott JS**. (2006). "*The structure of MbtI from Mycobacterium tuberculosis, the first enzyme in the biosynthesis of the siderophore mycobactin, reveals it to be a salicylate synthase*." *J Bacteriol*, **188**(17):6081-6091.
24. **Stegmann E, Rausch C, Stockert S, Burkert D, Wohlleben W**. (2006). "*The small MbtH-like protein encoded by an internal gene of the balhimycin biosynthetic gene cluster is not required for glycopeptide production*." *FEMS Microbiol Lett*, **262**(1):85-92.
25. **Wach MJ, Kers JA, Krasnoff SB, Loria R, Gibson DM**. (2005). "*Nitric oxide synthase inhibitors and nitric oxide donors modulate the biosynthesis of thaxtomin A, a nitrated phytotoxin produced by Streptomyces spp.*" *Nitric Oxide*, **12**(1):46-53.
26. **Buddha MR, Crane BR**. (2005). "*Structures of tryptophanyl-tRNA synthetase II from Deinococcus radiodurans bound to ATP and tryptophan. Insight into subunit cooperativity and domain motions linked to catalysis*." *J Biol Chem*, **280**(36):31965-31973.
27. **Buddha MR, Crane BR**. (2005). "*Structure and activity of an aminoacyl-tRNA synthetase that charges tRNA with nitro-tryptophan*." *Nat Struct Mol Biol*, **12**(3):274-275.
28. **Buddha MR, Keery KM, Crane BR**. (2004). "*An unusual tryptophanyl tRNA synthetase interacts with nitric oxide synthase in Deinococcus radiodurans*." *Proc Natl Acad Sci U S A*, **101**(45):15881-15886.
29. **Buddha MR, Tao T, Parry RJ, Crane BR**. (2004). "*Regioselective nitration of tryptophan by a complex between bacterial nitric-oxide synthase and tryptophanyl-tRNA synthetase*." *J Biol Chem*, **279**(48):49567-49570.
30. **Healy FG, Krasnoff SB, Wach M, Gibson DM, Loria R**. (2002). "*Involvement of a cytochrome P450 monooxygenase in thaxtomin A biosynthesis by Streptomyces acidiscabies*." *J Bacteriol*, **184**(7):2019-2029.

4.7 Web references

http://pir.georgetown.edu/pirwww/index.shtml - Protein Information Resource at Georgetown University Medical Center

CHAPTER 5

Antibiotics overproduction in *Streptomyces coelicolor* A3(2) mediated by phosphofructokinase A2 deletion

Irina Borodina[1]§, Jeroen Siebring[2]§, Jie Zhang[1], Colin Smith[3], Geertje van Keulen[4], Lubbert Dijkhuizen[2] and Jens Nielsen[1]

Center for Microbial Biotechnology, BioCentrum-DTU, 2800 Kgs. Lyngby, Denmark[1]; Department of Microbiology, Groningen Biotechnology and Biomolecular Science Institute (GBB), University of Groningen, PO Box 14, 9750 AA, Haren, The Netherlands[2]; Functional Genomics Laboratory, School of Biomedical and Molecular Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, UK[3]; Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK[4]

5.1 Abstract

Streptomycetes are exploited for production of a wide range of secondary metabolites and there is much interest in enhancing the level of production of these metabolites. Secondary metabolites are synthesized in dedicated biosynthetic routes, but precursors and co-factors are derived from the primary metabolism. High level production of antibiotics in Streptomycetes therefore requires engineering of the primary metabolism. Here we demonstrate this by targeting a key enzyme in glycolysis, phosphofructokinase, leading to improved antibiotic production in *Streptomyces coelicolor* A3(2). Deletion of *pfkA2* (SCO5426), one of three annotated *pfkA* homologues in *S. coelicolor* A3(2), resulted in a higher production of the pigmented antibiotics actinorhodin and undecylprodigiosin. The Δ*pfkA2* deletion strain had an increased carbon flux through the pentose phosphate pathway, as measured by ^{13}C metabolic flux analysis, establishing the ATP-dependent Pfk as a key player in determining the carbon flux distribution. The increased pentose phosphate pathway flux appeared largely due to accumulation of glucose-6-phosphate and fructose-6-phosphate, as experimentally observed in the mutant strain. Through genome-scale metabolic model simulations we predicted that decreased phosphofructokinase activity leads to increase in pentose phosphate pathway flux and in flux to pigmented antibiotics and pyruvate supposedly through increased production of NADPH and acetyl-CoA. Integrated analysis of gene expression data using the genome-scale metabolic model further revealed transcriptional changes in redox co-factors - dependent enzymes as well as in pentose phosphate pathway enzymes and enzymes involved in storage carbohydrates biosynthesis.

5.2 Introduction

Streptomyces are well known for their production of biologically active secondary metabolites, with almost two-thirds of all known natural antibiotics being produced by *Streptomyces*. The best genetically characterized strain is *S. coelicolor* A3(2), which produces pigmented antibiotics and has become the preferred model organism in *Streptomyces* research.

Secondary metabolites are generally synthesized in dedicated biosynthetic pathways, but these are linked to the primary metabolism through the use of specific precursors and various cofactors. Although still largely unclear, the link between primary and secondary metabolism has been repeatedly observed (1-3). Thus, deletion of the glyceraldehyde-3-phosphate (G3P) dehydrogenase encoding gene *gap1* improved clavulanic acid production in *S. clavuligerus* due to increased supply of precursor G3P (4). Furthermore, overexpression of acetyl-CoA carboxylase *acc* in *S. coelicolor* improved the yield of actinorhodin, which uses the ACC product, malonyl-CoA, as a precursor for actinorhodin biosynthesis (5). In another study deletion of important tricarboxylic acid cycle genes, such as citrate synthase (*citA*) or aconitase (*acoA*) in *S. coelicolor*, resulted in glutamate auxotrophy, overproduction of organic acids, as well as changes in secondary metabolite production and morphological differentiation (6,7). Polyphosphate kinase *ppk* gene inactivation in *S. lividans* was also shown to have effect on secondary metabolite production as it resulted in accumulation of polyphosphates and activation of actinorhodin production, which is normally silenced in this species (8-10).

In order to further evaluate the link between primary and secondary metabolism we decided to study the effect of phosphofructokinase (Pfk) on secondary metabolism in *S. coelicolor*. Pfk is a highly regulated enzyme in the conserved Embden-Meyerhof-Parnas glycolysis pathway (EMP) catalyzing the phosphoryl transfer to the 1st position of the substrate fructose-6-phosphate. This enzyme is most commonly dependent on ATP as the source of the phosphoryl group and is allosterically regulated by a number of effectors. The reaction catalyzed by ATP-dependent Pfks is irreversible and therefore the six-carbon sugars are confined to the subsequent glycolytic reactions. This positions Pfk as a committed step for the EMP pathway at glucose-6-phosphate (G6P) branching node. Besides going to EMP pathway the G6P can enter pentose phosphate pathway or serve as precursor for carbohydrate polymers biosynthesis. Alteration of phosphofructokinase activity might influence the flux distribution at the G6P node and hence it is a high profile target for altering metabolic fluxes in the cell.

The full genome sequence of *S. coelicolor* A3(2) (11) revealed the presence of three highly homologous *pfkA* genes (SCO2119, SCO5426 and SCO1214) with 56-72% identity and 69-82% similarity at the amino acid level. This number of copies is large when compared to closely related, less developmentally complex actinomycetes like *Mycobacterium tuberculosis* (12,13) and *Corynebacterium glutamicum* (14), which both contain only one *pfkA* gene. The biological reason for the presence of multiple copies of primary metabolic genes such as *pfk* is not known, and there is little information available about the regulation of primary metabolic isoenzymes in *S. coelicolor*. It is for instance not known whether expression of various isoenzymes is dependent on the developmental phase of a colony, or whether isoenzymes convey different kinetics.

Previously, a *S. coelicolor* Pfk enzyme had been purified and characterized as described by Alves et al. (15). This ATP-dependent, allosterically controlled enzyme was linked to the *pfkA1* gene (SCO2119). Studying the original N-terminal sequence of the protein purified in that study, however, revealed that the N-terminal sequence does not exactly fit the predicted PfkA1 N-terminus but has an identical match to the PfkA2 (SCO5426) enzyme N-terminus (Table 5-1). In this paper, we also report that PfkA2 is the only isoenzyme with detectable Pfk activity in *S. coelicolor* extracts (see below), which further confirms that the enzyme studied by Alves et al. is indeed PfkA2 (15) (pers. comm.).

Table 5-1. Alignment of N-terminal sequences of all three *S. coelicolor* PFK isoenzymes and the N-terminal sequence of the previously purified *S. coelicolor* PFK as determined by Alves (15).

```
PFKA1        MKVGVLTGGGDCPGLNAVIRAVVRKGVQEYGYDFTGFRDGWRGPLEGDTVPLDIPAVRGI
N terminus   MRIGVLTAGGDXPGLNAVIXSVVXXAVDENNXGXXV-----------------------
PFKA2        MRIGVLTAGGDCPGLNAVIRSVVHRAVDNYGDEVIGFEDGYAGLLDGRYRALDLNAVSGI
PFKA3        MRIGVLTSGGDCPGLNAVIRSVVHRAVVDHGDEVIGFRDGWKGLLECDYLKLDLDAVGGI
             *  **** *** *******  **   *
```

Changing the flux distribution at the principal glucose-6-phosphate (G6P) node by modulating Pfk expression might be a tool to influence antibiotic production through increasing cofactor availability. Thus, shifting the flux towards the pentose phosphate pathway (PPP) may increase the NADPH production. Since the biosynthetic pathways for many polyketides require NADPH, this may result in enhanced biosynthesis of antibiotics.

To evaluate whether PfkA2 in fact is controlling the flux distribution and whether there is a direct link between the fluxes in the primary metabolism and secondary

metabolism, we studied the effect of *pfkA2* deletion on intracellular carbon fluxes, metabolite levels, gene expression, and secondary metabolite production in *S. coelicolor*.

5.3 Experimental procedures

5.3.1 Alignment

Construction of the alignment of N-terminal sequences of all three PFK isoenzymes and the N-terminus as determined by Alves et al. (15) was done using ClustalX.

5.3.2 Strains

The *S. coelicolor* M145 (SCP1-, SCP2-) strain used for construction of the *pfkA* single gene deletions was kindly supplied by Matthias Redenbach. The ordered cosmid library made from this strain (16) was used in the *S. coelicolor* genome sequencing project (11). Later genotyping of the strain indicated a possible duplication of 20 genes (SCO4672-4695) mostly coding for proteins of unknown function. No gene was found underrepresented in this strain. The reference strain used for metabolic flux analysis was *S. coelicolor* M145, which was a gift from Mervyn Bibb, John Innes Centre, UK. To ensure that the unintended duplication in the Redenbach parental strain did not influence the observed phenotype, we also have deleted *pfkA2* in the reference M145 strain. No morphological differences between M145 and the Redenbach strain, or between the *pfkA2* mutants thereof, were observed on plates or in liquid complex medium. Ten genes from the duplicated region, most of them assigned as hypothetical proteins, were found to be overexpressed in the Redenbach Δ*pfkA2* mutant as compared to the reference M145.

The single gene deletion mutants Δ*pfkA1*, Δ*pfkA2* and Δ*pfkA3* were constructed by lambda-mediated recombination in *E. coli* according to the REDIRECT protocol (17). The REDIRECT strategy is based on replacement of a chromosomal sequence within an *S. coelicolor* cosmid (16) by a selectable marker. This marker is PCR generated using primers with 39 nt homology extensions. These homology extensions are designed to match the immediate flanking regions of the genes targeted allowing stable in frame, full length gene replacements.

The *pfkA1* gene was fully replaced with a viomycin resistance cassette, which was amplified from plasmid pIJ780. The *pfkA2* and *pfkA3* genes were replaced by an apramycin resistance cassette generated by PCR from plasmid pIJ773. The primers used in this process are listed in Table 5-2. The strains were maintained on master plates of mannitol soy flour (MS) agar (18) containing 50 μg/ml apramycin/viomycin for the mutants. The cells from the master plates were streaked onto fresh MS plates, from which spores were harvested to make dense spore plates. The dense spore plates provided inocula for further studies.

Table 5-2. Nucleotide sequence of primers used for λRED–mediated deletion of the *pfkA* genes. The universal 5' and 3' ends of the disruption cassette are underlined.

REDIRECT primers	sequence (5' – 3')
PfkA1 forward	GGCCTCAACCACCCACCCTGGCAAAGGGG TTGTGCGATGATTCCGGGGATCCGTCGACC
PfkA1 reverse	ACCGCCGTCGCCGGTCAGCGGCCGTTCCG CCGGACGTCATGTAGGCTGGAGCTGCTTC
PfkA2 forward	TCACCGGTGTCAACAGCGAGCAGGAGACA CAGCACGATGATTCCGGGGATCCGTCGACC
PfkA2 reverse	TGGACGGCGGTCGAGAGGCACCCCTAGAG GGGAGGCCTATGTAGGCTGGAGCTGCTTC
PfkA3 forward	CCGGACCCGTGCCACGTTGAGAAGGTATTT CGGGCTATGATTCCGGGGATCCGTCGACC
PfkA3 reverse	GCGAATCACCTCAAGGCCCGAGCTCCGACC GGCAGGTCATGTAGGCTGGAGCTGCTTC

5.3.3 Batch cultivations for metabolic flux analysis and transcription analysis

Batch cultivations were performed in fermenters with 1 L working volume (Applicon, USA) at 30ºC, pH 6.8-7.0, 800 rpm agitation rate, and a 1 vvm aeration rate. The fermentors were equipped with a cooled condenser to avoid evaporation from the medium. pH was controlled by automatic addition of 1 N NaOH. The aeration rate was adjusted after every sampling to compensate for working volume reduction. The concentrations of CO_2 and O_2 were monitored by an acoustic gas analyzer (1311, Bruël & Kjær, Denmark).

The defined minimal medium was limited in phosphate and contained 3 mM NaH_2PO_4, 100 mM NH_4Cl, 10 mM KCl, 2 mM Na_2SO_4, 2 mM citric acid as chelating agent, 1.25 mM $MgCl_2$, 0.25 mM $CaCl_2$, as well as per liter: 20 g glucose, 5 ml trace elements solution (20 mM $FeCl_3$, 10 mM $CuCl_2$, 50 mM $ZnCl_2$, 10 mM $MnCl_2$, 0.02 mM Na_2MoO_4, 20 mM $CoCl_2$, 10 mM H_3BO_4), 1 ml vitamins solution (0.05 g biotin, 1 g Ca-pantothenate, 1 g nicotinic acid, 25 g myo-inositol, 1 g thiamine-HCl, 1 g pyridoxine-HCl and 0.2 g para-aminobenzoic acid per liter), and 100 µl organic antifoam 204 (Sigma-Aldrich, USA). All components, except glucose and vitamins, were added to the MilliQ water (Millipore, USA) in the fermenter and autoclaved at 121ºC for 40 min. Glucose and vitamins were added after autoclaving and pH adjustment by sterile filtration. 30% of the added glucose for the labeled cultivations consisted of 1-^{13}C-glucose (Omicron Biochemicals Inc., USA).

The inoculum for each fermenter was prepared as follows: spores from one dense spore plate were harvested with 2 ml 20% sterile glycerol, filtered through glass fiber wool to get rid of mycelial fragments and inoculated into a 500 ml volume baffled shake flask with 50 ml 2xYT medium (18) and 30 glass beads (ø3 mm). The culture was incubated at 30°C and 150 rpm agitation for 12 h, centrifuged at 3,000 g for 5 min, resuspended in 3-5 ml of residual liquid and immediately inoculated into a fermenter.

5.3.4 Analysis of biomass dry weight

Throughout cultivation 12 ml samples were taken. Of these samples 8 ml was filtered through pre-dried Supor®-450 membrane filters (0.45 µm pore, Pall Corporation, USA). The filtrate was collected in a pre-cooled tube and kept at -20ºC until extracellular metabolite analysis. The cells were washed twice with 0.9% sodium chloride, dried in a microwave oven at 150 W for 20 min after which the dry weight of

the biomass was calculated. The remaining 4 ml of the sample was divided into 2x2 ml portions and kept at -20ºC for antibiotic analysis.

5.3.5 Shake flasks cultivations for analysis of antibiotics, intracellular sugar phosphates and carbohydrates

The defined mineral medium for the shake flasks cultivation was the same as described for the batch cultivations, but contained 40 g per liter of glucose instead of 20 g per liter, and was supplemented with 50 mM 2-(N-morpholino)ethanesulfonic acid for buffering purposes. The pH of the medium was adjusted to 6.8 before autoclaving.
The cultivations were performed in 500 ml baffled shake flasks with 50 ml defined medium and 30 glass beads (ø3 mm). The cultures were kept at 30ºC on a shaking table with an agitation speed of 150 rpm.
The inoculum for shake flasks was prepared in the same way as for the fermenters, except for the centrifugation step. Preculture (2.5 ml) was added directly to each shake flask (5% v/v).

5.3.6 Shake flask cultivations for antibiotic phenotyping in complex medium

Streptomyces strains used were grown under standard conditions in 200 ml R2YE medium (18). The 1 L flasks were inoculated with 10^6 spores per ml of medium. Spores were pregerminated for 15 min at 50°C in TY medium (18) and spun down (15 min, 16,100xg). Most of the supernatant was removed before inoculation. The cultures were grown in 1 L shake flasks equipped with stainless steel spirals at 30°C while shaken at 200 rpm.

5.3.7 Actinorhodin and undecylprodigiosin quantifications

To extract actinorhodin, 2 ml 3 N KOH was added to a 2 ml sample, vortexed and left shaking overnight at +2ºC. The suspensions were centrifuged at 4,000xg for 10 min and the absorbance of the supernatant was measured at 640 nm. The actinorhodin concentration was calculated based on an extinction coefficient of 25,320 $M^{-1}cm^{-1}$ (19). Undecylprodigiosin was extracted in the same manner but with methanol acidified to pH 1.5 with HCl in a tube with ø0.25-0.5 mm glass beads. The extinction coefficient of undecylprodigiosin used was 100,500 $M^{-1}cm^{-1}$ at 530 nm (20).

5.3.8 Analysis of extracellular metabolites

A range of extracellular metabolites including glucose were analyzed by HPLC using an Aminex HPX-87H column (Bio-Rad Laboratories, USA). The operating temperature was 60°C. Separation was achieved using 5 mM H_2SO_4 at a flow rate of 0.6 ml/min. The HPLC was equipped with a Waters 410 differential refractometer (Millipore, USA) and a tunable absorbance detector set at 210 nm (Waters 486, Millipore, USA). Glucose, glycerol, succinate, and ethanol were quantified using a refractometer whereas acetate and pyruvate were quantified by UV absorbance. A calibration curve was made using 6 different metabolite concentrations.

5.3.9 Analysis of intracellular sugar phosphates concentration

Cells from triplicate shake flask cultures were quenched in cold methanol and then extracted with chlorophorm (pH 7), according to a slightly modified method from de Koning and van Dam (21). Cell culture (10 ml) was used to determine biomass dry weight in order to calculate the sugar phosphate portion per g dry weight. Cell

suspension (5 ml) was sprayed into 20 ml 60% (v/v) methanol buffered with 12.5 mM tricine (pH 7.4) and kept in a -40ºC ethanol bath. The cells were centrifuged for 5 min at 4,000xg at -20 ºC. The supernatant was collected and kept at -80ºC until further analysis. Pure methanol (2.5 ml) was added to the cell pellet, after which the tube was immersed in liquid nitrogen and stored at -80ºC. The extraction was performed by shaking the tubes at -20ºC overnight at 300 rev/min after addition of 5 ml chlorophorm, 2 ml of buffer (3 mM tricine, 3 mM EDTA, pH 7.0), and glass beads ø0.25-0.5 mm. The mixture was centrifuged at 1,200xg for 12 min at -20ºC and the upper phase was collected. Pure methanol (2 ml) was added to the remaining chlorophorm and cells, vortexed, separated by centrifugation, and pooled with the first portion of the extract. The extract and the quenching liquid were freeze-dried and the solid matter was redissolved in 1 ml MilliQ water.
Glucose-6-phosphate (G6P) and fructose-6-phosphate (F6P) were analyzed in an enzymatic assay with G6P-dehydrogenase (G6P-DH) and phosphoglucose isomerase (Pgi) by fluorimetric measurement of NADPH formation on Perkin Elmer Luminescence spectrometer LS50B set at 340 nm for excitation and 456 nm for emission. The reaction was carried out at room temperature in 250 mM triethanolamine buffer with 0.025 mM $NADP^+$. The sample was added to the buffer and the baseline absorbance was measured after which 1 unit of G6P-DH was added per ml. The increase in absorbance was measured after 5-7 min. The increase in absorbance, and thus NADPH, is proportional to the conversion of G6P. The additional increase in absorbance after addition of Pgi corresponds to the amount of F6P converted.
G6P and F6P amounts were calculated from the calibration curves, which were linear in the 0 to 10.5 nM range. The samples fell within the range of the calibration curve. Some of the samples were measured a few times to determine the measurement error, which was not higher than 5%.

5.3.10 Measurement of carbohydrates content

The samples for measurement of carbohydrate content were collected during the stationary phase when the triplicate shake flask cultures became red indicating undecylprodigiosin production. Sample (5-10 ml) was centrifuged at 10,000xg for 10 min, washed twice with distilled water and freeze-dried. 5 mg of dried sample was then dispersed in 1 ml distilled water. To this mixture 2 ml of 0.5% anthrone in 70% sulfuric acid was added (22). The sample was boiled in a water bath for 20 min, cooled to room temperature, and absorbance was measured at 625 nm and compared to a glucose standard curve.

5.3.11 Measurement of amino acids labeling and estimation of metabolic fluxes

For the labeled cultivations additional samples of 5-10 ml were withdrawn. The biomass was centrifuged at 10,000xg for 5 min, washed twice with cold 0.9% NaCl and kept at -20ºC until further analysis. The biomass was hydrolyzed with 6 N HCl at 105ºC and then derivatized as described in Christensen et al. (23). The labeling of amino acid fragments was measured by GC-MS (23). The labeling of each fragment was expressed as summed fractional labeling (SFL) determined by the formula:

$$SFL = \frac{i_0 \cdot 0 + i_1 \cdot 1 + i_2 \cdot 2 + \ldots + i_n \cdot n}{i_0 + i_1 + i_2 + \ldots + i_n},$$

where i_n is the peak intensity of mass isotopomer with n ^{13}C atoms. The data was corrected for natural isotope abundance before usage (23).
The labeling data together with growth parameters and cellular composition were used to calculate fluxes by an algorithm from Wiechert (24) implemented in an in-house software package based on Matlab (The MathWorks Inc., USA). The cellular composition was taken from Borodina et al. (25), except for protein composition, which was specifically determined for *S. coelicolor* cultivations on a defined mineral medium (26) (Appendix 5.1).

5.3.12 Diamide sensitivity assays

Diamide sensitivity assays were performed as described previously (3,27). Halo surfaces were determined after 48 h of incubation at 30℃.

5.3.13 Sampling for microarray analysis and RNA isolation

The samples for gene expression analysis were taken during the exponential growth phase, when the carbon dioxide concentration in the outlet gas was 0.1%, corresponding to approximately 2 g/L of dry weight biomass. Samples of 20 ml were withdrawn quickly and sprayed into 50 ml falcon tubes filled with crushed ice resulting in almost immediate cooling down to 0-2 ºC. The mixture was shaken and centrifuged at 4,000 g for 1 min. The supernatant was decanted, the cells were immersed in liquid nitrogen and then stored at -80ºC until RNA extraction. The total time spent on the sampling procedure did not exceed two minutes.
Total RNA extraction was performed with an RNeasy kit (QIAGEN GmbH, Germany) using a FastPrep FP120 (3x40", Qbiogene Inc., USA) for breaking the cells. After the removal of cellular debris additional steps were performed: two rounds of phenol:chlorophorm:isoamylalcohol (pH 8) extraction and 1 round of chlorophorm extraction. The rest of the procedure was according to the manufacturer's instructions. The total RNA was subjected to DNase digestion and then concentrated from the reaction solution using an RNeasy CleanUp kit (QIAGEN GmbH, Germany). The quality of RNA was assessed on a Bioanalyzer (Agilent, USA) to ensure that no degradation of 23S RNA of 16S RNA had occurred.

5.3.14 Microarray analysis

The arrays were spotted oligonucleotide arrays (University of Surrey fabricated oligonucleotide (50 mer) microarrays), containing duplicate spots of probes for 7337 *S. coelicolor* genes. Each probe was 50 nucleotide bases long and was designed for minimal cross-hybridization with other genes.
The RNA samples were labeled and hybridized to the arrays according to the procedure described on http://www.surrey.ac.uk/SBMS/Fgenomics/Microarrays. The design included three biological replicates for the reference strain and two biological replicates for the Δ*pfkA2* mutant as well as dye swaps. Only two biological replicates were analyzed in case of Δ*pfkA2* because one of the RNA preparations did not pass the quality control. We applied loop design (28), where samples were compared directly to each other without a common reference (Appendix 5.2).

5.3.15 Array image analysis and data normalization

The arrays were scanned using an Affymetrix 428 laser scanner (Affymetrix Inc., USA). Image analyses were carried out using BlueFuse software (BlueGnome Limited, UK), which allows automatic grid alignment, spot boundary detection,

background substraction, and signal intensities estimation. The quality of the spots was scored 0 or 1 (0 for bad quality, 1 for good) depending on spots radius, circularity, uniformity, and grid offset. The raw data can be found in Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE7686.
For the data analysis we used MAANOVA program (http://www.jax.org/staff/churchill/labsite/software/Rmaanova) implemented in R statistical package (http://www.r-project.org). The raw signal intensities from two spots corresponding to each gene were log2 transformed, normalized within arrays using regional lowess method, which is a combination of intensity-based and special lowess methods (29). The normalized data was fitted into ANOVA model. The ANOVA model was a fixed effect linear model with variance terms due to array, dye and strain effects (30). The differentially expressed genes were found by permutation test (30). Only genes that had at least one quality spot per each array were considered.

5.3.16 Identifying reporter metabolites and significantly co-regulated metabolic Subnetworks

An algorithm by Patil *et al.* (31) and a *S. coelicolor* genome-scale metabolic model (25) were used for identifying reporter metabolites and transcriptionally co-regulated metabolic subnetworks.

5.3.17 Metabolic modeling

A genome-scale metabolic model of *S. coelicolor* A3(2) containing 700 unique reactions was used for metabolic modeling (25). For calculating fluxes in the wild type strain the glucose uptake rate was set to the experimental value and the fluxes through $NADP^+$-dependent glucose-6-phosphate 1-dehydrogenase (G6P-DH), phosphofructokinase, isocitrate dehydrogenase, and through the reactions leading from pyruvate to acetyl-CoA, were constrained to the values determined from ^{13}C metabolic flux analysis (MFA) for the wild type strain. The flux distribution was obtained by linear optimization for growth using the LINDO program (Lindo Systems Inc., USA). To simulate the fluxes in the mutants the flux through phosphofructokinase was set to decreasing values and the new flux distribution was found by using the MOMA approach (32) using in-house developed software (33).

5.3.18 Reverse Transcriptase-PCR

RNA for RT-PCR was isolated according to van Keulen *et al*. (34) in which the usage of Tri®Reagent (Sigma) is combined with the usage of the SV total RNA isolation kit of Promega. RNA (5 µg) was mixed with 0.5 µl of 10x hexanucleotide mix (Roche), RNase-free DNaseI (Promega) and MilliQ until a final volume of 12 µl. This mixture was then heated to 70°C for 10 min after which 4 µl 5x first strand buffer for SuperScriptII (Invitrogen), 2 µl 0.1 M DTT, and 1 µl 10mM dNTP mix (Roche) were added. This mix was incubated for 10 min at 25°C before adding 1 µl RNaseOut (10 U/µl, Invitrogen) and 1 µl SuperScript™II Reverse Transcriptase (200 U/µl, Invitrogen). During the incubation period at 42°C (1 h) cDNA is made of the total RNA batch.
The PCR on cDNA template was performed with the primers listed in Table 5-3. The internal fragments amplified are in the range of 520-530 bp. The PCR mix contained 10-100 ng cDNA, 1 pmol/µl of both forward and reverse primer, 2.5 mM $MgCl_2$, 5%

dimethylsulfoxide (DMSO) and 1 U of Taq polymerase (Sigma). The annealing temperature was 56.5°C and 40 cycles were applied.

Table 5-3. Nucleotide sequences of primers used for amplifying internal *pfkA* fragments in RT-PCR.

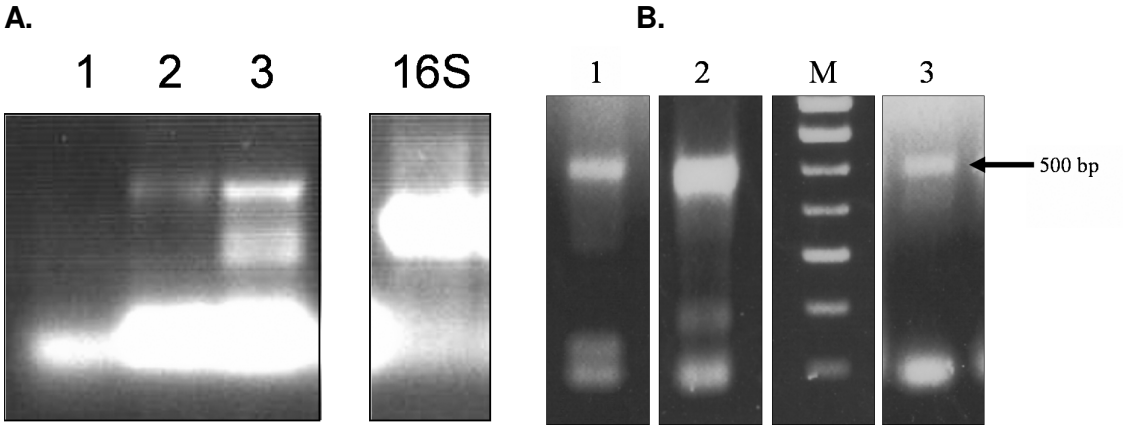
RT primer	Sequence (5' – 3')
RT PFK1 forward	GACGGCATCCGGCGCATCAA
RT PFK1 reverse	TCTGCTTGGCCAGCCATTCG
RT PFK2 forward	GCGAGGCCTGCGAGAACGCG
RT PFK2 reverse	AGCGCGGTGCCGATGCCCTG
RT PFK3 forward	GGACGGCGTGGAGCGGGCCC
RT PFK3 reverse	AGCTGGCGGGCGATCCCGGC

Restriction analysis on the RT-PCR products was performed following standard procedures in order to determine the authenticity of the products. High similarity of the *pfkA* genes causes cross annealing of the primers producing aspecific RT-PCR products. To verify identity of *pfkA1*, *pfkA2*, or *pfkA3* specific products, the PCR product was cut, respectively with *Sac*II, *Bgl*I, or *Hae*II (Figures 5-1 and 5-2). As a control, a restriction was performed with any two of the aforementioned restriction enzymes. The RT-PCR products obtained were purified and concentrated prior to digestion using a Sigma GenElute™ PCR Clean-Up Kit.

Figure 5-1. Schematic representation of the restriction digestion procedure used for confirming *pfkA* products in Reverse Transcriptase-PCR. Presence of recognition sites in the *pfkA* internal fragments for any of the three restriction endonucleases mentioned at the left of the scheme is symbolized with an X.

	pfkA1	*pfkA2*	*pfkA3*
Sac II	———	X	X
Bgl I	X	———	X
Hae II	X	X	———

Figure 5-2. A. RT-PCR data on *S. coelicolor* A3(2) reference strain cDNA template, created from total mRNA isolated. From left to right RT-PCR products are shown that were obtained using *pfkA1, pfkA2, pfkA3,* and 16S primers. **B.** Restriction fragments of PCR products obtained by RT-PCR. Lane 1 represents *pfkA2* specific PCR product amplified from the *S. coelicolor* A3(2) reference strain cDNA pool of which any false positive signals have been removed by restriction digestion (see figure 5-1). Lane 2 shows confirmed positive *pfkA3* signal obtained from reference strain cDNA. Lane 3 is *pfkA3* product obtained from Δ*pfkA2* cDNA and confirmed by restriction analysis. M is the marker.



5.3.19 Preparation of extracts and enzyme activity assays

Cells used for activity assays were harvested after 67 h of growth in 200 ml defined medium in a 1 L flask equipped with a stainless steel spiral. Cells were harvested by centrifugation and resuspended in buffer containing 50 mM TES [*N*-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid] (pH 7.2), 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol, 50 mM $(NH_4)_2SO_4$, and 0.1 mM phenylmethylsulfonyl fluoride (buffer A), and disrupted by sonication until >90% of the cells were lysed.

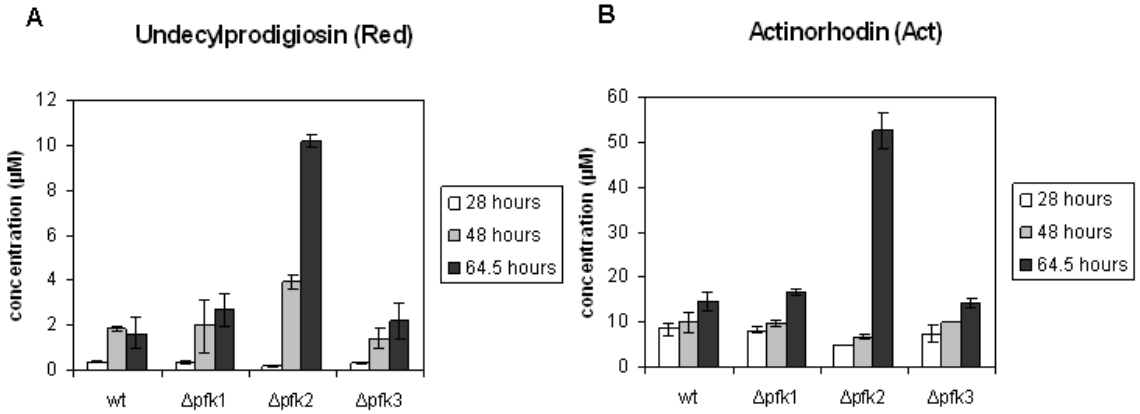
Pfk activity assays were based on the consumption of NADH in a coupled assay, according to Alves *et al.* (1997) (15). G6P-DH (EC 1.1.1.49) assays were based on the production of NADPH and performed according to the protocol of Lessie and Vander Wyk and modified by Butler *et al*. (3). Both the consumption of NADH and the production of NADPH were measured spectrophotometrically at 340 nm. Protein concentrations were determined using the Bio-Rad protein assay reagent and a BSA based standard curve.

5.4 Results

5.4.1 Deletion of phosphofructokinase isoenzymes influences antibiotic production

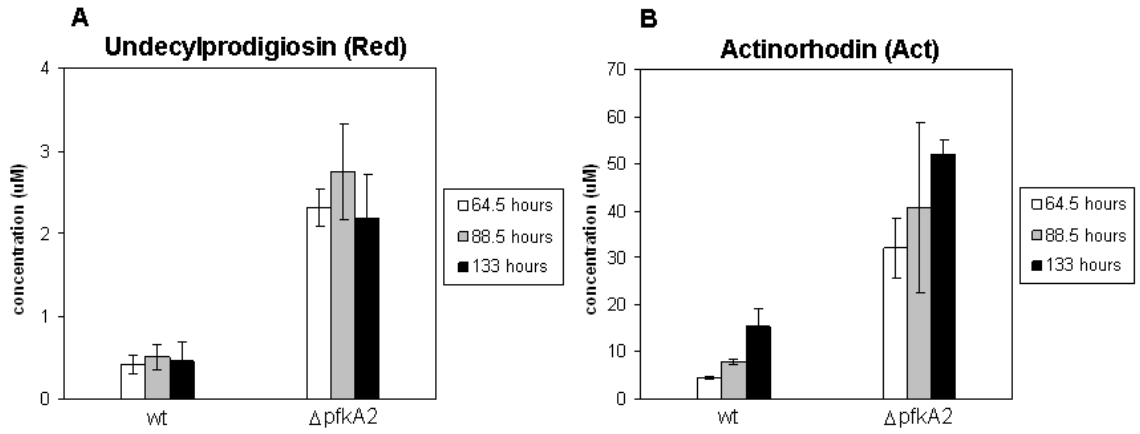
The *S. coelicolor* reference strain as well as its isogenic mutants Δ*pfkA*, Δ*pfkA2* and Δ*pfkA3*, were grown in complex (R2YE) media in shake flasks and the pigmented antibiotics actinorhodin (ACT) and undecylprodigiosin (RED) were quantified in the course of several days (Figure 5-3).

Figure 5-3. Quantification of undecylprodigiosin (Red, A) and actinorhodin (Act, B) in R2YE grown *pfkA* deletion cultures. Concentrations of Red and Act produced were determined after 28, 48, and 64.5 hours of growth in shake flasks corresponding to exponential, early stationary and late stationary phases.



Morphology of the *pfkA* single deletion mutants did not differ from that of the parent strains, neither in liquid media nor on solid media. When grown in R2YE liquid medium the Δ*pfkA2* mutant had a 2-6 fold increased antibiotic production compared to the reference strain. This was in contrast with the *pfkA1* and *pfkA3* single deletion mutants, which produced ACT and RED in quantities comparable to that of the reference strain. The ACT and RED overproduction by the *pfkA2* deletion strain was most pronounced after 64.5 h of growth when the culture entered the stationary phase. The ACT and RED overproduction phenotype in Δ*pfkA2* was also observed in defined medium (Figure 5-4). Because of its antibiotic overproduction characteristics, the Δ*pfkA2* strain was selected for more detailed analysis.

Figure 5-4. Antibiotics production by reference and Δ*pfkA2* mutant strains during growth in shake flasks in defined medium. The error bars indicate standard deviation between biological triplicates.



5.4.2 Deletion of pfkA2 influences biomass yield

The reference M145 strain and the Δ*pfkA2* mutant were grown in phosphate-limited defined mineral medium with glucose as the only carbon source in well-controlled

fermenters to estimate growth parameters (Table 5-4). The maximal specific growth rates were not significantly different when accounting for the standard deviation (0.09±0.02 h^{-1} and 0.11±0.01 h^{-1}), however, the mutant had a lower biomass yield on glucose (0.31±0.02 against 0.39±0.02 g [dry weight]/g glucose). The mutant strain was also characterized by a slightly larger specific carbon dioxide production, 54±7 mmol CO_2/g (dry weight) compared to 44±7 mmol CO_2/g (dry weight) in the reference strain (93% confidence level). HPLC analysis of the fermentation broth of the Δ*pfkA2* cultures revealed the presence of small amounts of acetate, which was not found in the reference cultivations.

Table 5-4. Growth parameters of reference and Δ*pfkA2* mutant strains as determined in controlled fermentations in defined medium. The values for the reference strain are average values from three cultivations and the values for the mutant are averaged from duplicate cultivations.

Growth parameters	**M145****	**M145Δ*pfkA2******
Maximal specific growth rate, μ_{max}, h^{-1}	0.11±0.01	0.09±0.02
Biomass yield on glucose, Y_{SX}, g (dry weight)·g $glucose^{-1}$	0.39±0.02	0.31±0.02
Specific glucose uptake rate, r_s, mmol glucose·g (dry weight)$^{-1}$·h^{-1}	1.57	1.61
CO_2 yield on glucose, Y_{SC}, mmol CO_2·g (dry weight)$^{-1}$	44 ±7	54 ±7
Carbon balance closure*	94±9%	84±4%

* Assuming 5% ash content in the biomass and average biomass composition $CH_{1.81}O_{0.58}N_{0.2}S_{0.004}P_{0.01}$
** Average values and standard deviations calculated from 4 independent cultivations
*** Average values and standard deviations calculated from 3 independent cultivations

5.4.3 The ΔpfkA2 strain accumulates glucose-6-phosphate

We speculated that limited phosphofructokinase capacity in the Δ*pfkA2* strain resulted in intracellular accumulation of the substrate for this enzyme - fructose-6-phosphate. As the isomerisation between F6P and G6P is generally close to equilibrium, we expected to see elevated levels of G6P as well. Both strains were grown in triplicate shake flasks and the samples for intracellular sugar phosphate analysis were taken 24 h after inoculation. At this point the biomass concentration was about 2 g/L and the cells were in the exponential growth phase. Cold methanol quenching was applied. Because previous reports on actinomycetes (35) suggest leakage caused by cold shock, we measured the sugar phosphates both in extracts of biomass and in the quenching liquid. Indeed, we found that 94-97% of the metabolites had leaked out of the cells during the quenching with cold methanol. The method was, however, still suitable for our assay, because G6P and F6P do not freely diffuse across membranes and are normally not found outside cells; all sugar phosphates in the quenching liquid should therefore be of intracellular origin.

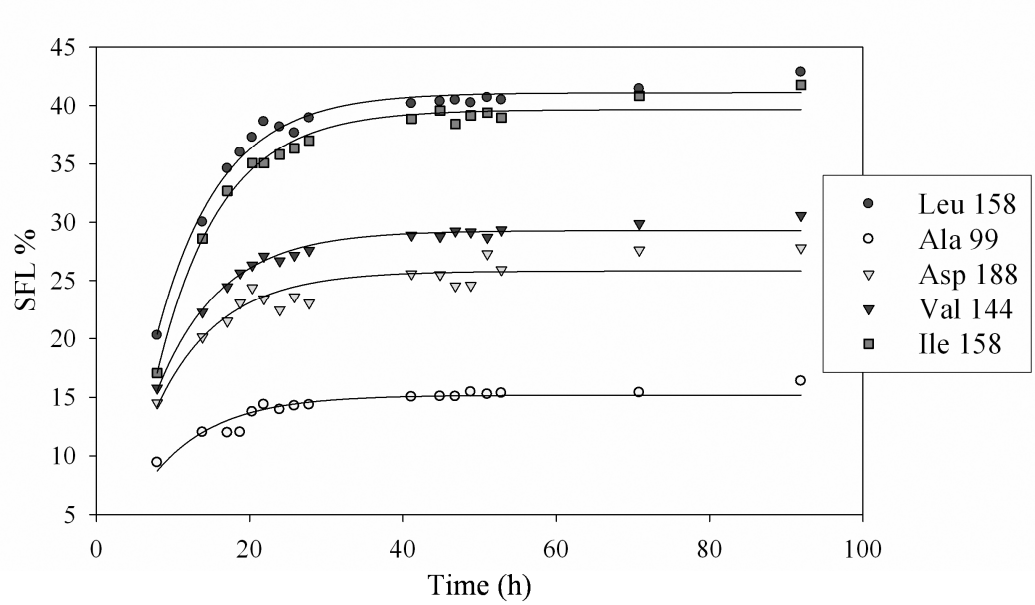
Glucose-6-phosphate concentration was higher in the mutant (6.9±0.2 μmol/g [dry weight]) than in the reference strain (5.2±0.4 μmol/g [dry weight]). Fructose-6-phopshate concentration was also higher at 90% confidence level: 1.9±0.4 μmol/g [dry weight] for the mutant and 1.4±0.4 μmol/g [dry weight] for the reference strain. The ratio between F6P- and G6P concentrations was 1:3.6 in both strains. The results were consistent with data in yeast, where deletions of one or two subunits of phosphofructokinase also resulted in accumulation of intracellular hexose monophosphates (36, 37). Another reported effect in yeast is the accumulation of storage carbohydrates glycogen and trehalose (37), carbohydrates that are also produced by *S. coelicolor* (38-40). We could not detect glycogen in the late stationary phase of the liquid cultures of either the reference or the mutant strain (data not

shown). The total carbohydrate content in stationary phase cultures was not significantly different between the two strains: 13±3% of dry weight for the reference strain and 15±2% for the mutant.

5.4.4 The ΔpfkA2 strain has an increased flux through the pentose phosphate pathway

Fermentations with the M145 and *ΔpfkA2* strains were performed with 30% 1-^{13}C-labeled glucose as the carbon source. Samples for biomass measurements were taken during cultivations and the labeling of proteinogenic amino acids was measured using gas chromatography coupled to mass spectrometry (GC-MS). Labeling of most of the amino acids fragments followed first order kinetics according to the equation: $SFL=SFL_0+(SFL_\infty-SFL_0)\times(1-e^{-\mu t})$ (Figure 5-5, Appendix 5.3), where SFL_0 stands for initial labeling of the fragment and SFL_∞ describes the steady state labeling after all the unlabeled proteins from the inoculum cells had been substituted by proteins synthesized from labeled glucose.

Figure 5-5. SFL values of leucine, alanine, aspartate, valine and isoleucine fragments as a function of time in the *S. coelicolor* A3(2) reference strain (the number next to the amino acid abbreviation is m/z value). The labeling of the amino acids fragments follows first order kinetics. Fermentations with the *S. coelicolor* A3(2) reference strain and the *ΔpfkA2* strain were performed with 30% 1-^{13}C-glucose as carbon source. Samples for biomass measurements were taken during cultivations and the labeling of proteinogenic amino acids was measured using gas chromatography coupled to mass spectrometry (GC-MS).



Labeling of most of the fragments had reached steady state by the end of the exponential phase. There were no clear changes in the labeling of amino acids after the end of the exponential growth phase, implying that flux distribution in the stationary phase was similar to that of the exponential phase. SFL_∞ values together with growth parameters were used to calculate the intracellular fluxes (Figure 5-6, Appendix 5.3).

The label distribution in the amino acids reflects the pathways by which amino acid precursors were synthesized. Labeling of alanine and valine gives information about the glycolytic pathway, which leads to the formation of pyruvate. With an active Entner-Doudoroff (ED) pathway present, half of the pyruvate made from 1-^{13}C-glucose would be labeled at the carboxylic group (1st position). The percentage of labeling of that atom was around 3% (deduced from fragments Ala158, Ala116, Ala99) and close to the natural labeling of 1.1% (Appendix 5.3). Also the 1st and 2nd atoms of phosphoenolpyruvate (fragment Phe143) were slightly labeled. Therefore we conclude that the label in pyruvate was not supplied by an ED pathway, but was the result of scrambling in the PP pathway and/or activity of phosphoenolpyruvate carboxykinase or malic enzyme, feeding label from the TCA cycle. In the oxidative branch of the pentose phosphate pathway the 1st carbon atom of glucose does not get incorporated in pyruvate but is oxidized, producing carbon dioxide. In case of glucose labeled in the C1 position, the label will be lost and all pyruvate resulting from this pathway will be non-labeled. The higher the flux through the PP pathway, the fewer ^{13}C heavy isotopes will get incorporated into pyruvate and into TCA cycle intermediates. In the reference *S. coelicolor* strain the labeling of pyruvate 2nd and 3rd carbon atoms at the end of cultivation (Ala99, Ala116) reached 16.4-16.6%, which is close to the theoretical maximum if all pyruvate is generated through the EMP pathway, meaning that the pentose phosphate flux was very low. For the mutant strain the corresponding labeling value was 14.2-14.5%, indicating higher PP flux. The labeling of pyruvate and TCA cycle intermediates (oxaloacetate and α-ketoglutarate) was lower in the Δ*pfkA2* strain, which also indicates a higher flux through the PP pathway in the mutant. The labeling of acetyl-CoA 2nd carbon atom as calculated from fragment Leu158 was lower than the labeling of the 3rd carbon atom than pyruvate. It means that not all acetyl-CoA is coming directly from pyruvate, but some can be supplied through other pathways that have lower labeling. It could be phosphoketolase pathway. Phosphoketolase Xpk (E.C. 4.1.2.9) catalyzes the cleavage of xylulose-5-phosphate into glyceraldehyde-3-phosphate and acetyl-phosphate as well as cleavage of fructose-6-phosphate into erythrose-4-phosphate and acetyl-phosphate. The activity of this enzyme has been measured mostly during growth on pentoses in *Bifidobacteria* (36,37), several *Lactobacillus* species (38,39), *Saccharomyces cerevisiae* (40) and others. BLASTing with Xpk protein sequence of *Bifidobacterium adolescentis* against the whole genome of *S. coelicolor* A3(2) gave a hit to a probable phosphoketolase SCO0617 with 49% identity. In the initial genome annotation this protein was noted as hypothetical (11).

For flux calculations we included the EMP and PP glycolytic pathways, phosphoketolase reactions and the tricarboxylic acid cycle. The anaplerotic glyoxylate shunt was considered to be inactive on glucose (41), so only pyruvate- and PEP carboxylases were included. It is not possible to discriminate between these two reactions with high confidence when the 1st carbon of pyruvate is not labeled, hence the summed anaplerotic flux is shown in Figure 5-6. We did not account for acetate secretion since only small amounts below 0.1 g/l were detected in the late broth samples of the mutant strain.

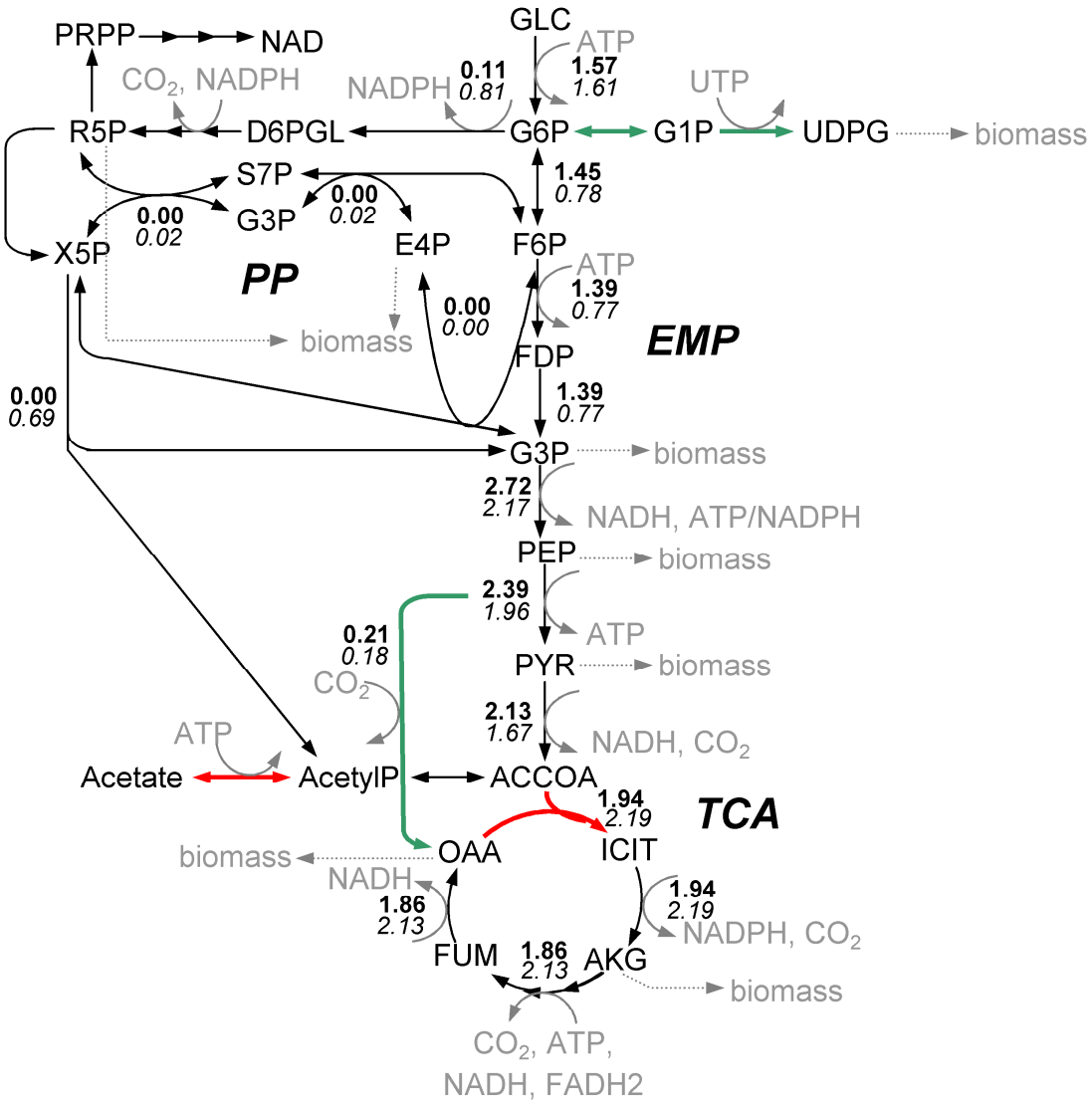
For *S. coelicolor* M145, 7% of the imported glucose went into the oxidative branch of the pentose phosphate (PP) pathway (Appendix 5.4). This pathway is involved in generation of precursors for nucleotide biosynthesis (ribose-5-phosphate) and aromatic amino acids (erythrose-4-phosphate). Because there was no back-flux from

the PP pathway into glycolysis, the main role of the PP pathway was generation of biomass precursors.

In the Δ*pfkA2* mutant strain as much as 50% of the imported glucose was shunted towards the pentose phosphate pathway. As the amount of precursors generated in this case exceeded the demand, most of the carbon entering the PP pathway was further shuffled through phosphoketolase into glyceraldehyde-3-phosphate and acetyl-phosphate. The oxidation of glucose-6-phosphate into ribulose-5-phosphate usually leads to production of two NADPH molecules. In *S. coelicolor* the cofactor specificity of two glucose-6-phosphate-1-dehydrogenases (*zwf1* and *2*, SCO6661 and SCO1937 respectively) and of 6-phosphogluconate dehydrogenase (*zwf3*, SCO0975) has been determined (5). Both enzymes are $NADP^+$-dependent which means that there was an increased NADPH production in the mutant strain. By the action of phosphate acetyltransferase acetyl-phosphate can be converted into acetyl-CoA, which further enters TCA cycle, fatty acid metabolism, secondary metabolism, etc. Acetyl-phosphate can also be hydrolyzed by acylphosphatase into acetate.

The flux through the TCA cycle and through the anaplerotic pathways did not differ much between the reference and the mutant strains.

Figure 5-6. Changes in intracellular fluxes in the Δ*pfkA2* mutant strain compared to the *S. coelicolor* A3(2) reference strain. The standard deviation of flux estimation did not exceed 5%. All fluxes are expressed in mmol/g(dry weight)/h. The enzymes that are part of significantly co-regulated metabolic subnetwork are highlighted. The whole subnetwork can be found in Appendix 5.6. Genes up-regulated in the mutant are shown in red and down-regulated in green. The reaction catalyzed by phosphoketolase and leading from fructose-6-phosphate into erythrose-4-phosphate and acetyl-phosphate is not shown to avoid crowdiness of the figure. The flux through it was 0.02 for the reference strain and 0.00 for the mutant strain.



5.4.5 The ΔpfkA2 strain produces more NADPH

To test whether the increased flux through the PP pathway in the mutant led to NADPH overproduction, we examined the diamide sensitivity of the strains. The test was previously used for a study of *zwf1/zwf2* negative *S. lividans* strains (3). Diamide oxidizes sulphide groups in proteins and thiols causing formation of toxic disulfide bridges. This oxidative stress is neutralized by thioredoxin reductase (27), which

requires NADPH as electron donor. The more NADPH is available in the cells, the more resistant the strain should be to diamide. The test showed a slightly reduced sensitivity of the mutant strain to diamide (at P>90%). The halo surface surrounding the 0.5 M diamide impregnated disc, measured 962.2 ± 27.5 mm^2 for the wild type versus 916.9 ± 15.5 mm^2 for the deletion strain.

5.4.6 Reverse transcriptase-PCR

High similarity of the three *pfkA* genes prevents discrimination between their transcripts on microarrays. This was also evident from our array studies, showing no significant change in *pfkA2* expression for the *pfkA2* deletion strain, indicating cross-hybridization by *pfkA1* and/or *pfkA3* mRNA (Appendix 5.5). A combination of RT-PCR with restriction digestion, however, does allow detection and discrimination of mRNA of a single *pfkA*. RT-PCR revealed the presence of *pfkA2* and *pfkA3* mRNA in M145 grown on mineral medium (Figure 5-2). The cells were harvested in the exponential growth phase. This finding is in accordance to that of Hesketh *et al*. (42) who detected PfkA2 and PfkA3 protein spots on 2D-gels of *S. coelicolor* grown under comparable conditions (SMM phosphate limited mineral medium). As anticipated, only the presence of *pfkA3* mRNA was detected in mRNA isolated from the *pfkA2* deletion strain harvested at the same time-point. The authenticity of the RT-PCR products was confirmed by restriction analysis, to exclude products formed due to cross-annealing of the primers.

5.4.7 Phosphofructokinase (Pfk) and glucose-6-phosphate 1-dehydrogenase (G6P-DH) activity assays

In order to study differences in G6P-DH and Pfk activities, enzyme assays were performed on *S. coelicolor* crude extracts. G6P-DH assays did not reveal significant differences in activity between *pfkA2* deletion mutants and their parent strain (166 ± 37 µmol/min/mg protein). Despite many attempts, no residual Pfk activity could be detected in *pfkA2* deletion mutants. Various assay conditions have been tested, aiming to detect even minute Pfk activities that might account for the remaining EMP carbon flow as detected by ^{13}C metabolic flux analysis, but with no success.

5.4.8 Deletion of pfkA2 causes transcriptional changes in regulatory and membrane proteins as well as in fatty acids metabolism

The effect of the *pfkA2* deletion on gene expression was studied by comparison of RNA of the *S. coelicolor* A3(2) reference- and mutant strains using spotted DNA microarrays. RNA was isolated from cultures in the mid exponential growth phase.
The 588 genes that had a p-value below 0.05 in permutation test and for which there was at least one quality spot per each array were considered to be differentially expressed (Appendix 5.5). The genes were sorted into functional categories according to a protein classification scheme proposed by the Sanger Institute (http://www.sanger.ac.uk/Projects/S_coelicolor/scheme.shtml). Functional categories that were most influenced by the genetic change were determined by hypergeometric distribution and included fatty acids metabolism, regulatory genes and membrane proteins (Table 5-5).

Table 5-5. Genes that are significantly (p<0.05) up- or down-regulated (dark shading) in the Δ*pfkA2* mutant in comparison to the *S. coelicolor* A3(2) reference strain. The genes are sorted by functional categories and for each category the p-value from hypergeometric distribution is shown.

Functional category	Differentially expressed	Total number in genome	p-value*
Transcriptional regulators and other regulatory proteins	48	460	0.01
Degradation of fatty acids	9	58	0.02
Gram positive membrane	72	826	0.09
Ribosomal proteins, synthesis, modification	6	62	0.20
Secondary metabolism	24	276	0.21
Gram positive peptidoglycan, teichoic acid	4	40	0.23
Cell division	3	20	0.27
Sensor kinase	7	86	0.33
Degradation of proteins, peptides, glycoproteins	10	126	0.36
Degradation of carbon compounds	7	90	0.38
Cations transport	3	39	0.39
Transposon/insertion element-related functions	4	58	0.46
Non-classified	42	559	0.49
Response regulators and sigma factors	12	167	0.51
Gram positive exported lipoprotein	32	438	0.55
Fatty acid and phosphatidic acid biosynthesis	3	53	0.60
Transport/binding proteins	36	518	0.69
Macromolecule synthesis, modification	16	247	0.70
Degradation of polysaccharides	5	92	0.71
Electron transport	3	70	0.79
Detoxification	2	88	0.97
Others	62	1017	-
Unknown function	178	2336	-
Total	588	7726	

*p-value of having the detected or higher number of genes from a particular functional category among the differentially expressed genes. Calculated according to: $p = 1 - \sum_{i=0}^{x} \frac{\binom{M}{i}\binom{N-M}{K-i}}{\binom{N}{K}}$, where N are the total genes analyzed on chip; K are differentially expressed genes; M are the genes from a particular functional category present on the chip and x is the number of differentially expressed genes that belong to that functional category. Small p-values indicate functional categories that were most affected by the mutation.

Of around 460 transcriptional regulators and regulatory proteins that were annotated in the *S. coelicolor* genome 46 changed their expression significantly in the Δ*pfkA2* mutant. As the exact role and binding sites are not known for most of the transcriptional regulators it is difficult to say what influence their changed expression has on the cells. One of them is, however, a better studied A-factor receptor homologue, CprB (SCO6071) (43). In contrast to another A-factor receptor homologue, CprA (SCO6312), which acts as a positive regulator of secondary

metabolism and sporulation, CprB seems to have an opposite role (44). Deletion of CprB caused increased actinorhodin production and earlier sporulation (44). This gene was down-regulated in Δ*pfkA2* mutant. Another gene with a better studied function was *recX* gene (SCO5770), which was slightly higher expressed in the mutant. *recX* is positioned downstream recombinase *recA* and is most probably involved in its downregulation (45-48). *recX* transcripts were previously detected in *S. lividans* only under DNA damaging conditions (49).
Fatty acids are building blocks of phospholipids and glycolipids used in the cellular membranes as well as of triacylglycerols, which serve as energy storage molecules (50-52). Triacylglycerols were found in the post-exponential phase biomass of several *Streptomyces* species and were hypothesized to serve as a C2 source for polyketides formation (50). Biosynthesis of fatty acids can as well serve as a sink for NADPH, which is used during chain elongation. Subsequent degradation of fatty acids would provide ATP and NADH/$FADH_2$, the latter also coupled to ATP production through the electron transport chain. In our data set we saw significant overexpression of acyl-CoA dehydrogenases (SCO3051, SCO4681, SCO1701) as well as lipases/esterases (SCO4746, SCO3644).

Interestingly, the phosphofructokinase genes display a trend of increased expression. Because of the high homology between the isoenzymes, it is not possible to distinguish the species on the array. However, since *pfkA2* has been completely deleted from the mutant, the mRNA must be of *pfkA1* and/or *pfkA3* origin, that in a reaction to the loss of *pfkA2* increase in expression. Because mRNA of *pfkA1* could not be detected with RT-PCR the mRNA must be the result of *pfkA3* transcription.

5.4.9 Metabolic subnetworks with changed gene expression

Of the 588 genes with altered expression only a minor part can be directly linked to the primary metabolism or to the antibiotic production. Studying metabolic subnetworks might in this case prove to be a useful tool for interpreting the effect of the *pfkA2* deletion on gene expression. Metabolic genes are interconnected through common metabolites. To avoid potentially harmful accumulation of metabolites, the activities of the enzymes that produce and consume those metabolites have to be coordinated. In bacteria, genes that belong to a linear metabolic pathway often are positioned in one operon to ensure their coordinated expression. When an enzyme is deleted or over-expressed, the cells will seek to maintain metabolic *status quo* by changing the activities of connected enzymes, either at the transcriptional level or at the protein level.
To further analyze the transcriptional response to the deletion of *pfkA2* we identified metabolites around which the most significant changes in expression occurred (Table 5-6). For this purpose we used a genome-scale metabolic model to define connections between metabolites and the enzymes that have these metabolites as substrates or products (31). The significance of the gene expression test, expressed as p-values from the permutation test, were used for each metabolic gene included in the model.

Table 5-6. Top 20 reporter metabolites* around which the most significant changes in expression occurred in the Δ*pfkA2* mutant strain compared to the *S. coelicolor* A3(2) reference strain, identified using an algorithm by Patil *et al.* (31) and an *S. coelicolor* genome-scale metabolic model (25), to define connections between metabolites and the enzymes that have these metabolites as substrates or products.

Metabolite	Processes	Number of neighbors	p-value
D-Glucose 1-phosphate	Storage carbohydrates	2	0.00
D-Galactose 1-phosphate	Storage carbohydrates	2	0.07
UTP	Pyrimidines/Storage carbohydrates/Peptidoglycan	5	0.02
N6-(1,2-Dicarboxyethyl)-AMP	ATP biosynthesis	2	0.05
NADH	Reducing cofactor	49	0.04
D-Ribose 5-phosphate	Pentose phosphate pathway	9	0.08
2,3,4,5-Tetrahydrodipicolinate	Amino acids metabolism	2	0.04
L-Lysine	Amino acids metabolism	3	0.07
3-Phosphonooxypyruvate	Amino acids metabolism	2	0.00
L-Homoserine	Amino acids metabolism	2	0.00
L-Histidine	Amino acids metabolism	3	0.02
L-Aspartate	Amino acids metabolism	12	0.03
L-Aspartate 4-semialdehyde	Amino acids metabolism	4	0.03
2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate	Amino acids metabolism	2	0.07
Acetaldehyde	Pyruvate metabolism	7	0.01
Ethanol	Pyruvate metabolism	3	0.02
Propionyl-phosphate	Propanoate metabolism	2	0.02
Propanoate	Propanoate metabolism	4	0.04
Methylmalonate semialdehyde	Propanoate metabolism	7	0.08
Oxaloacetate	TCA cycle/Anaplerotic reactions	11	0.03

Among the high scoring reporter metabolites were glucose-1-phoshate and galactose-1-phosphate, precursors for storage carbohydrates, which indicates transcriptional changes in this route. Increased accumulation of storage carbohydrates was observed in yeast with knocked-out phosphofructokinase (53), though we could not detect that in *S. coelicolor* (see above). Another top scoring reporter metabolite is the reducing co-factor NADH. This cofactor participates in 62 reactions in the model and the low p-value indicates that transcriptional changes occurred in many of the neighboring enzymes. Some other reporter metabolites were from amino acid biosynthesis and degradation pathways. Amino acid biosynthesis leads to oxidation of NADPH into $NADP^+$ and can be connected to the altered level of this co-factor. There were changes in pyruvate and propanoate metabolism, which supply precursors for polyketide compounds. The presence of oxaloacetate among reporter metabolites can indicate both transcriptional changes in the TCA cycle and anaplerotic reactions. Remarkably, significant transcriptional changes occurred around ribose-5-phosphate, a key metabolite of the pentose phosphate pathway and nucleotide precursor.

The enzymes can be represented in the form of a network, where the enzymes act as nodes and the shared substrates make the edges. In this metabolic network significantly regulated sub-networks can be found (31). Genes linked to the most significant sub-networks are highlighted in Figure 5-6. The sub-network

included citrate synthase *citA*, which is homologous to the principal isoenzyme in the *S. coelicolor* strain MT1110 (7). The strain with deletion of this enzyme becomes auxotrophic for glutamate and shows severe acidification of medium with acetate, pyruvate and other organic acids (7). *CitA* (SCO2736) was up-regulated in our mutant strain (p=0.08). Another up-regulated enzyme, which is also part of the most significantly regulated subnetwork was acetate kinase (SCO5424). The analysis indicates some transcriptional changes around the acetyl-CoA node, which is also consistent with the low acetate secretion by the mutant strain. The subnetwork also included RNA synthesis, which is primarily catalyzed by alpha- (SCO4729), beta- (SCO4654) and omega- (SCO1478) RNA polymerase subunits. In the mutant, alpha-, beta- and omega-chains of RNA polymerase were downregulated, the former two at a 95%- and the latter at an 85% significance level.

5.4.10 Metabolic modeling

For the purpose of metabolic engineering it is a great advantage to be able to predict which genetic changes will lead to a desired phenotype. Genome-scale metabolic models have been used recently for making such predictions (54,55). We were interested to see what phenotype the existing *S. coelicolor* genome-scale metabolic model would suggest for decreased phosphofructokinase activity.

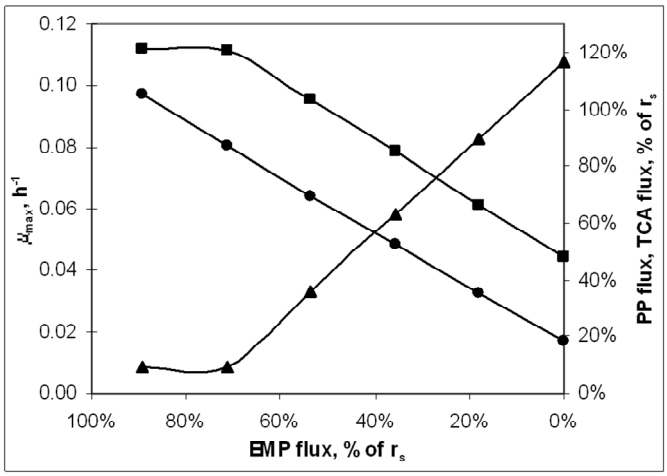
The model contains 700 unique reactions, many of which are reversible, hence the degree of freedom is very high and the model can therefore not be used to obtain precise estimates of the intracellular fluxes without the addition of some constraints. By including a few constraints from the ^{13}C MFA calculations (restrictions of the PP flux, the TCA flux, and the EMP flux), the model could predict the specific growth rate to be 0.1 h^{-1} when the specific glucose uptake rate was set at 1.57 mmol/g (dry weight)/h. This is very close to the actual value (Table 5-4) and hence we assumed that the predicted fluxes were a good reflection of the actual fluxes in the cell.

Using the obtained wild type flux distribution we predicted how the fluxes would change with a decrease in the EMP flux. We assumed that the cells will undergo as few flux changes as possible to adjust to the new conditions, hence we chose a minimization of flux adjustment (MOMA) approach for the simulations (32).

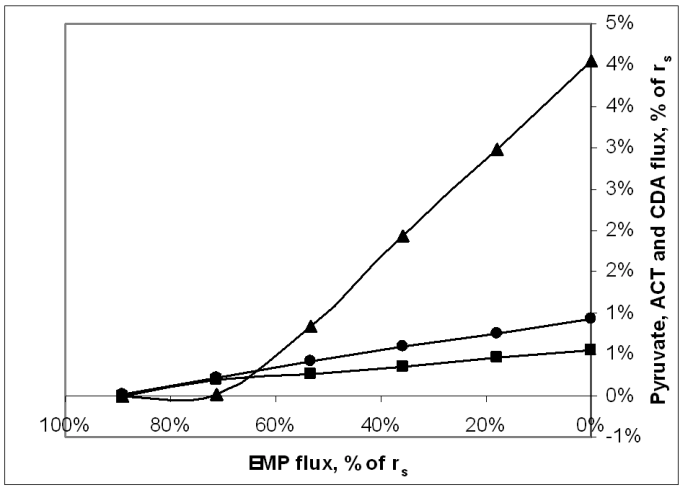
First, the model predicted excretion of compounds, which were not detected experimentally and are not earlier reported to be produced during *S. coelicolor* fermentations, e.g. acetate, acetaldehyde, ethanol, formate, and proline. When the secretion of these compounds was constrained to zero, the model predicted pyruvate, actinorhodin, and calcium-dependent antibiotic as fermentation products (Figure 5-7). The decrease of flux through phosphofructokinase correlated with decreased specific growth rate and an increased flux through the PP pathway, the latter was also observed experimentally.

Figure 5-7. Influence of decreased flux through phosphofructokinase on growth rate (●), pentose phosphate flux (▲) and TCA cycle flux (■) [A] and on pyruvate excretion (▲), actinorhodin (■) flux and calcium-dependent antibiotic flux (●) [B].

A



B



5.5 Discussion

In this study the effects of the deletion of *pfkA2* (SCO5426) on cellular processes in *S. coelicolor* were studied. The most obvious effect of this deletion was an increase in production of the pigmented antibiotics actinorhodin (ACT) and undecylprodigiosin (RED) on both mineral and rich media without a significant decrease in specific growth rate.

A direct link between primary- and secondary metabolism in Streptomycetes has been demonstrated before. In *Streptomyces lividans*, a close relative of *S. coelicolor*, deletion of 1 of 2 *zwf* genes led to an increase in ACT and RED production with an unaltered specific growth rate (3). The *zwf* gene products, glucose-6-phosphate dehydrogenase (G6P-DH), catalyze the first reaction of the pentose phosphate

pathway. The authors explained their results by increased carbon availability for antibiotics biosynthesis pathway when less carbon was oxidized to carbon dioxide in pentose phosphate pathway. Another group performed metabolic flux analysis on the physiological data from *S. lividans* chemostats and also hypothesized that decreased flux through the PP pathway would result in increased antibiotic production (56). This hypothesis was tested in *S. coelicolor* by Ryu *et al* (5), who partially blocked the flow into the PPP by deleting *zwf* and increased actinorhodin precursor availability by overexpressing acetyl-CoA carboxylase. Deleting the main Zwf activity (*zwf2*), however, caused an enhanced mycelial growth and a decreased specific ACT production rate. These results were in contradiction to the ones obtained for *S. lividans* (3). In both *S. lividans* and *S. coelicolor* the deletion of a *zwf* gene did lead to a decrease in G6P-DH specific activity. However, it should be noted that *in vivo* fluxes do not obligatory correlate with the *in vitro* measured enzymatic activity, therefore it is not clear how the flux through PPP was influenced in these studies.

The current study focused not on decreasing C-flow through the PPP but on decreasing flux through glycolysis. Glycolysis is not completely blocked in the Δ*pfkA2* mutant, as demonstrated by the metabolic flux analyses, despite our inability to find remaining Pfk activity in Δ*pfkA2* cell extracts. The remaining Pfk activity is most likely provided by PfkA3, of which mRNA could be traced using Reverse Transcriptase PCR. Failure to detect this PfkA3 activity may be due to enzyme instability during extract preparation, or requirement for specific assay conditions that we have not been able to identify. This reduction in Pfk activity in the mutant led to an accumulation of F6P and G6P. This result is consistent with data in yeast, where deletions of one or two subunits of phosphofructokinase also resulted in accumulation of intracellular hexose monophosphates (57). Our studies show that accumulation of G6P in *S. coelicolor* A3(2) results in an increased flow of carbon through the PPP, which is consistent with the increased CO_2 production by the Δ*pfkA2* mutant strain. Also a flux through xylulose-5-phosphate phosphoketolase was predicted in the Δ*pfkA2* mutant, but not in the reference strain. This enzyme is usually active during growth on pentoses, which give plenty of xylulose-5-phosphate that the cells have to deal with. Apparently, similar effect was achieved by increasing the glucose conversion through the pentose phosphate pathway. Through the labeling pattern talks for the presence of phosphoketolase activity, it should additionally be proven in enzymatic essays. It is interesting to note that over-expression of PPP enzymes would not necessarily lead to higher PPP flux. The higher affinity of phosphofructokinase to fructose-6-phosphate may cause it to be the flux controlling for the PPP flow. The final result of over-expressing PP pathway related enzymes might therefore only cause a minor increase in the PPP flux. Our study indicates that due to the tight regulation in the central carbon metabolism, it may be better to constrain fluxes in the metabolic network through deletion of specific flux controlling genes rather than by over-expression of genes.

We suggest that the stimulating effect of the increased PPP flux on antibiotic production is at least partly due to an increase in supply of NADPH as demonstrated by diamide sensitivity tests. NADPH is the reducing agent used in the process of making secondary metabolites and the PPP is one of the most important NADPH producing pathways. A role for NADPH in increase of antibiotic production has also been suggested before (1). Another effect could be increased production of acetyl-CoA, which serves as the basic precursor for polyketides, due to high flux thorough

the phosphoketolase pathway. The later hypothesis, however, needs more verification.
The various observations made can be explained in terms of metabolite overflow. Accumulation of primary metabolites can be potentially harmful to an organism. Excess metabolites can be disposed of through the production of waste products, possibly the original function of the relatively ancient secondary metabolic pathways. The major evolutionary driving force for secondary metabolite synthesis, however, may soon have become that of fitness increase due to their bioactivity (58). Nevertheless, the overflow concept still gives a foundation for speculation on the link between primary- and secondary metabolism. In light of this hypothesis any gross metabolic imbalance that cannot be dealt with by tweaking primary metabolic pathways would stimulate production of secondary metabolites.
We also show that the experimentally observed effects of decreased phosphofructokinase activity can be predicted using a genome-scale metabolic model. This implies that the model can be used for design of strains with further improved antibiotic production.
Gene expression analysis, which revealed numerous changes in the expression of regulatory proteins in the mutant, indicate that increase of antibiotics production also might be influenced by some regulatory events, e.g. that *pfkA2* deletion caused higher expression of the antibiotics biosynthetic clusters. As we did not analyze genes expression in the stationary phase, we can not confirm or decline this hypothesis, however, we could see that the indirect negative regulator of actinorhodin biosynthesis CprB was down-regulated in the mutant in the exponential growth phase.
Transcriptional data yet also showed changes in the expression of metabolic genes. The total number of genes that were found significantly differentially expressed was relatively low, less than 10% of the genome. This could be expected considering that the mutation did not have a drastic effect on cellular physiology and morphological development. Interpretation of gene expression data is even further complicated by the inferior annotation of *S. coelicolor* genes as compared to more studied model organisms like *E. coli* or the yeast *Saccharomyces cerevisiae*. Integrated data analysis methods have proven useful for interpretation of –omic data. For studying transcriptional changes in metabolic genes particularly supplementing the expression data analysis with the knowledge of metabolic network has an advantage (31). We identified transcriptional changes in the carbohydrate storage metabolism, as well as in enzymes connected to reducing co-factor NADH, PP pathway and nucleotides metabolism intermediate ribose-5-phosphate and tricarboxylic cycle intermediates oxaloacetate and acetyl-CoA. These findings confirm that changes have occurred in the central carbon metabolism and resulted in higher antibiotics biosynthesis.
In future work it will be of interest to study whether the mutant with deletion of *pfkA2* can be useful for production of other NADPH- and/or acetyl-CoA-requiring antibiotics.

FOOTNOTES

* We thank Mervyn Bibb (John Innes Centre, UK) and Matthias Redenbach (Technical University Kaiserslautern, Germany) for providing us with *S. coelicolor* M145, Michael Jewett (Technical University of Denmark, Denmark) for advice and help with intracellular metabolites extraction, Hans Peter Smits (Fluxome Science, Denmark) for advice and help with sugar phosphates analysis. IB thanks Technical University of Denmark for financial support through Ph.D. scholarship.

§ These authors contributed equally to the work.

This research was sponsored in part by the EU 6FP ActinoGEN project (Contract LSHM-CT-2004-005224).

5.6 References

1. Gunnarsson, N., Eliasson, A. and Nielsen, J. (2004) *Adv Biochem Eng Biotechnol* **88**, 137-178
2. Bibb, M. J. (2005) *Curr Opin Microbiol* **8**, 208-215
3. Butler, M. J., Bruheim, P., Jovetic, S., Marinelli, F., Postma, P. W. and Bibb, M. J. (2002) *Appl Environ Microbiol* **68**, 4731-4739
4. Li, R. and Townsend, C. A. (2006) *Metab Eng* **8**, 240-252
5. Ryu, Y. G., Butler, M. J., Chater, K. F. and Lee, K. J. (2006) *Appl Environ Microbiol*
6. Viollier, P. H., Nguyen, K. T., Minas, W., Folcher, M., Dale, G. E. and Thompson, C. J. (2001) *J Bacteriol* **183**, 3193-3203
7. Viollier, P. H., Minas, W., Dale, G. E., Folcher, M. and Thompson, C. J. (2001) *J Bacteriol* **183**, 3184-3192
8. Chouayekh, H. and Virolle, M. J. (2002) *Mol Microbiol* **43**, 919-930
9. Ghorbel, S., Smirnov, A., Chouayekh, H., Sperandio, B., Esnault, C., Kormanec, J. and Virolle, M. J. (2006) *J Bacteriol* **188**, 6269-6276
10. Ghorbel, S., Kormanec, J., Artus, A. and Virolle, M. J. (2006) *J Bacteriol* **188**, 677-686
11. Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A. and Goble, A. (2002) *Nature* **417**, 141-147
12. Cole, S. T. and Barrell, B. G. (1998) *Novartis Found Symp* **217**, 160-72; discussion 172-7
13. Camus, J. C., Pryor, M. J., Medigue, C. and Cole, S. T. (2002) *Microbiology* **148**, 2967-2973
14. Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, B. J., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Kramer, R., Linke, B., McHardy, A. C., Meyer, F., Mockel, B. and Pfefferle (2003) *J Biotechnol* **104**, 5-25
15. Alves, A. M., Euverink, G. J., Bibb, M. J. and Dijkhuizen, L. (1997) *Appl Environ Microbiol* **63**, 956-961
16. Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D. A. (1996) *Mol Microbiol* **21**, 77-96
17. Gust, B., Challis, G. L., Fowler, K., Kieser, T. and Chater, K. F. (2003) *Proc Natl Acad Sci U S A* **100**, 1541-1546
18. Kieser, T., Bibb, M., Buttner, M., Chater, K. and D.A., H. (2000) *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich
19. Bystrykh, L. V., Fernandez-Moreno, M. A., Herrema, J. K., Malpartida, F., Hopwood, D. A. and Dijkhuizen, L. (1996) *J Bacteriol* **178**, 2238-2244
20. Tsao, S. W., Rudd, B. A., He, X. G., Chang, C. J. and Floss, H. G. (1985) *J Antibiot* **38**, 128-131
21. de Koning, W. and van Dam, K. (1992) *Anal Biochem* **204**, 118-123

22. Morris, D. L. (1948) *Science* **107**, 254-255
23. Christensen, B. and Nielsen, J. (1999) *Metab Eng* **1**, 282-290
24. Wiechert, W. (2001) *Metab Eng* **3**, 195-206
25. Borodina, I., Krabben, P. and Nielsen, J. (2005) *Genome Res* **15**, 820-829
26. Davidson, A. (1992) *Quantitative Microbial Physiology of Streptomyces coelicolor* A3(2), PhD thesis, University of Glasgow, Glasgow
27. Paget, M. S., Kang, J. G., Roe, J. H. and Buttner, M. J. (1998) *EMBO J* **17**, 5776-5782
28. Vinciotti, V., Khanin, R., D'Alimonte, D., Liu, X., Cattini, N., Hotchkiss, G., Bucca, G., O, d. J., Rasaiyaah, J., Smith, C. P., Kellam, P. and Wit, E. (2005) *Bioinformatics* **21**, 492-501
29. Yang, Y. H., Dudoit, S., Luu, P. and Speed, T. (2001) in *Microarrays: optical technologies and informatics* (Bittner,M., Chen,Y., Dorsel,A. and Dougherty, E., eds.), pp. 141-152, SPIE, San Jose, CA, USA
30. Kerr, M. K., Martin, M. and Churchill, G. A. (2000) *J Comput Biol* **7**, 819-837
31. Patil, K. R. and Nielsen, J. (2005) *Proc Natl Acad Sci U S A* **102**, 2685-2689
32. Segre, D., Vitkup, D. and Church, G. M. (2002) *Proc Natl Acad Sci U S A* **99**, 15112-15117
33. Patil, K. R., Rocha, I., Forster, J. and Nielsen, J. (2005) *BMC Bioinformatics* **6**, 308
34. van Keulen, G., Siebring, J., Rembacz, K. P., Hoogeveen, M., Tomczynska, M. and Dijkhuizen, L. (2004) *J Microbiol Methods* **58**, 139-142
35. Wittmann, C., Kromer, J. O., Kiefer, P., Binz, T. and Heinzle, E. (2004) *Anal Biochem* **327**, 135-139
36. Meile, L., Rohr, L. M., Geissmann, T. A., Herensperger, M. and Teuber, M. (2001) *J Bacteriol* **183**, 2929-2936
37. Orban, J. I. and Patterson, J. A. (2000) *J Microbiol Methods* **40**, 221-224
38. Heath, E. C., Hurwitz, J., Horecher, B. L. and Ginsburg, A. (1958) *J Biol Chem* **231**, 1009-1029
39. Lampen, J. O., Gest, H. and Sowden, J. C. (1951) *J Bacteriol* **61**, 97-98
40. Sonderegger, M., Schumperli, M. and Sauer, U. (2004) *Appl Environ Microbiol* **70**, 2892-2897
41. Rollin, C., Morgant, V., Guyonvarch, A. and Guerquin-Kern, J. L. (1995) *Eur J Biochem / FEBS* **227**, 488-493
42. Hesketh, A. R., Chandra, G., Shaw, A. D., Rowland, J. J., Kell, D. B., Bibb, M. J. and Chater, K. F. (2002) *Mol Microbiol* **46**, 917-932
43. Natsume, R., Takeshita, R., Sugiyama, M., Ohnishi, Y., Senda, T. and Horinouchi, S. (2003) *Acta Crystallogr D Biol Crystallogr* **59**, 2313-2315
44. Onaka, H., Nakagawa, T. and Horinouchi, S. (1998) *Mol Microbiol* **28**, 743-753
45. Stohl, E. A., Brockman, J. P., Burkle, K. L., Morimatsu, K., Kowalczykowski, S. C. and Seifert, H. S. (2003) *J Biol Chem* **278**, 2278-2285
46. Sheng, D., Liu, R., Xu, Z., Singh, P., Shen, B. and Hua, Y. (2005) *DNA Repair (Amst)* **4**, 671-678
47. Pages, V., Koffel-Schwartz, N. and Fuchs, R. P. (2003) *DNA Repair (Amst)* **2**, 273-284
48. Galvao, C. W., Pedrosa, F. O., Souza, E. M., Yates, M. G., Chubatsu, L. S. and Steffens, M. B. (2003) *Can J Microbiol* **49**, 145-150
49. Vierling, S., Weber, T., Wohlleben, W. and Muth, G. (2000) *J Bacteriol* **182**, 4005-4011

50. Olukoshi, E. R. and Packter, N. M. (1994) *Microbiology* **140**, 931-943
51. Olukoshi, E. R. and Packter, N. M. (1992) *Biochem Soc Trans* **20**, 99
52. Packter, N. M. and Olukoshi, E. R. (1995) *Arch Microbiol* **164**, 420-427
53. Wilson, W. A., Wang, Z. and Roach, P. J. (2002) *Mol Cell Proteomics* **1**, 232-242
54. Alper, H., Jin, Y. S., Moxley, J. F. and Stephanopoulos, G. (2005) *Metab Eng* **7**, 155-164
55. Alper, H., Miyaoku, K. and Stephanopoulos, G. (2005) *Nat Biotechnol* **23**, 612-616
56. Avignone Rossa, White, J., Kuiper, A., Postma, P. W., Bibb, M. and Teixeira de Mattos, M. J. (2002) *Metab Eng* **4**, 138-150
57. Heinisch, J. (1986) *Curr Genet* **11**, 227-234
58. Vining, L. C. (1992) *Gene* **115**, 135-140

CHAPTER 6

Role of global regulator ppGpp for gene expression in liquid cultures of *Streptomyces coelicolor* A3(2)

Irina Borodina[1], Colin Smith[2], Mervyn Bibb[3] and Jens Nielsen[1]

[1] Center for Microbial Biotechnology, BioCentrum-DTU, 2800 Kgs. Lyngby, Denmark
[2] Functional Genomics Laboratory, School of Biomedical and Molecular Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, UK
[3] Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, UK

6.1 Abstract

In industrial processes antibiotic-producing *Streptomyces* are usually grown in batch cultivations on complex medium, containing amino acids and peptides as nitrogen source. Exhaustion of nitrogen source elicits ppGpp-mediated stringent response, which is known to cause growth cessation and induction of antibiotics biosynthesis genes in the model *Streptomyces* – *S. coelicolor* A3(2).
We compared growth, amino acids consumption and gene expression in *S. coelicolor* and its congenic Δ*relA* relaxed mutant grown in liquid cultures on nitrogen-limited medium in order to elucidate the global role of ppGpp.
Deletion of *relA* gene caused the mutant to grow with a higher specific growth rate but with a lower biomass yield on glucose and amino acids. Amino acids in both reference and mutant strains were taken up simultaneously and the uptake rates were proportional to the extracellular concentrations of amino acids, pointing that the main uptake routes were through non-specific amino acids transporters. The largest transcriptional changes occurred at the peak of ppGpp concentration between the end of the exponential and the early stationary phase in the reference strain, while the transcriptional changes in the mutant strain at the corresponding physiological stages were moderate. The ppGpp-regulated genes included ribosome constituents, secondary metabolic genes, and also cellular membrane constituents and numerous regulatory genes. The upregulated genes were overrepresented on the arms of the chromosome, while downregulated genes were more common on the core of the chromosome around the origin of replication. Analysis of significantly co-regulated metabolic networks identified transcriptional changes in energy metabolism in the stringent strain in the transition phase. The corresponding changes were not found in the relaxed mutant.

KEYWORDS

Stringent response, *Streptomyces coelicolor*, gene expression

ABBREVIATIONS

ppGpp	guanosine 3´,5´-bispyrophosphate
pppGpp	guanosine 3´-diphosphate, 5´-triphosphate
RNAP	RNA polymerase
SHX	serine hydroxamate

6.2 Introduction

Under the conditions of nutrient starvation microorganisms can elicit a stringent response characterized by slow-down of growth, reduction of stable RNAs synthesis as well as changes in the morphology, pathogenicity or secondary metabolites production [Sands et al, 1952]. The response is mediated by guanosine 3´,5´-bispyrophosphate (ppGpp), synthesized by stringent factor (p)ppGpp synthase RelA, which catalyses ppGpp synthesis either directly from GDP and ATP or from GTP and ATP through guanosine 3´-diphosphate, 5´-triphosphate (pppGpp) [Cashel et al, 1969]. The enzyme is associated with the ribosomal complex and apparently gets activated when free tRNA is encountered at the A-site of 50S ribosome [Haseltine et al, 1973]. The presence of free tRNA can be a consequence of nitrogen or carbon starvation, but can as well be provoked by sudden oxygen depletion, which leads to

reduced ATP synthesis and hence abolishment of the ATP-dependent transamination of free tRNAs.
At growth-favorable conditions constant low levels of ppGpp and pppGpp are maintained by the combined action of a synthase and a hydrolase, which in *E. coli* are coded by respectively *relA* and *spoT* and in actinomycetes are typically fused in one gene.

The presence of ppGpp-mediated stringent response was demonstrated in several *Streptomyces* spp. and its influence on morphological differentiation and secondary metabolites formation was shown [Oki et al, 1975; An et al, 1978; Riesenberg et al, 1984; Ochi et al, 1986]. Specifically, the *relC* relaxed mutants, which had a mutation in the ribosome, preventing the binding of uncharged amino acids and hence activation of the stringent factor, were deficient in antibiotics production.

In *S. coelicolor* a decrease in transcription from the four promoters of the *rrnD* gene set coding ribosomal RNA was measured when stringent response was induced by addition of serine hydroxamate, which acts as a competitive inhibitor of seryl-tRNA synthase [Strauch et al, 1991]. Strauch et al. have also shown that when *S. coelicolor* A3(2) is grown in a nitrogen-limited medium, the accumulation of ppGpp peaks at the end of exponential phase and precedes entry into stationary phase and antibiotics biosynthesis. They detected that this time point also correlated with the appearance of transcripts of an early gene in the actinorhodin biosynthetic cluster - *actIII*. Chakraburtty et al. continued the work on stringent response in *S. coelicolor* A3(2). They isolated and sequenced and then tried to disrupt the *relA* gene. Deletion of a C-terminal fragment did not result in a ppGpp-deficient strain [Chakraburtty et al, 1996], however, the deletion of a bigger N-terminal region on the *relA* gene gave a mutant that did not synthesize ppGpp under any of the conditions tested [Chakraburtty et al, 1997]. It is most likely that the C-terminal region of *relA* is responsible for the ribosome binding, while N-terminal region has the catalytic activity and on deletion of the C-terminal region the protein still can function in the cells in a ribosome-independent manner like the stringent factor in *S. marookaensis* [Oki et al, 1975; Muta et al, 1992]. The mutant showed relaxed phenotype where transcription from the *rrnD* promoters did not decrease significantly at the nutritional shift-down. On the nitrogen-limited medium the mutant was deficient in morphological differentiation as well as antibiotics (undecylprodigiosin and actinorhodin) production. On phosphate-limited medium the mutant sporulated and produced antibiotics normally, meaning that phosphate-limitation is regulated in a different, *relA*-independent manner [Chakraburtty et al, 1997].

In order to study the direct ppGpp effect on antibiotics genes, the *relA* gene and its truncated versions were placed under the control of a thiostrepton-inducible promoter (*tipA*p) [Hesketh et al, 2001]. Addition of thiostrepton to the exponentially growing cultures lead to the accumulation of intracellular ppGpp and induced the pathway specific activator for actinorhodin production (*actII-ORF4*), but not the activator for undecylprodigiosin (*redD*). The growth was not influenced by *relA* induction and the production of undecylprodigiosin started upon the entry into stationary phase as in the wild type. It was also previously described by Takano and co-workers that provoking the ppGpp formation by nutritional shift-down or by serine hydroxamate addition does not lead to *redD* induction [Takano et al, 1992]. Apparently the

presence of ppGpp is not sufficient to activate the transcription of *redD*, which might be additionally controlled by some other regulatory genes or γ-butyrolactones, shown to appear in the *S. coelicolor* A3(2) culture broth during the transition phase [Aigle et al, 2000].

Over the past ten years the evidence has been accumulated that ppGpp physically interacts with RNA polymerase (RNAP) in *E. coli* [Reddy et al, 1995]. The binding site is constituted of three RNAP subunits: β, β´ and σ^{70} [Hernandez et al, 1995; Chatterji et al, 1998; Zhou et al, 1998; Toulokhonov et al, 2001; Cashel et al, 2003]. It's been explicitly proven by Jishage et al. that ppGpp modulates the competition of different σ-factors for the core RNA polymerase [Jishage et al, 2002]. The accumulation of ppGpp leads to dissociation of the core RNAP from stringent promoters as *rrn* and the increased concentration of free cores allows promoters with lower affinity for σ^{70} and promoters that use alternative σ-factors to recruit the polymerase and thus start the transcription of the controlled regulon. One type of the regulon that are positively regulated by ppGpp is amino acids biosynthesis clusters, proven by the fact that $ppGpp^0$ *E. coli* strains are polyauxotrophic [Xiao et al, 1991]. The transcription from alternative promoters controlled by σ^S and σ^{32} in *E. coli* also positively correlates with ppGpp presence and is drastically reduced in $ppGpp^0$ cells. The mechanism discovered in *E. coli* is probably universal for all the organisms that exhibit stringent response. For *Streptomyces* spp., which have unusually high amount of σ-factors (65 in *S. coelicolor* and 62 in *S. avermitilis*) [Bentley et al, 2002; Omura et al, 2001], the ppGpp can play a very important global regulatory role changing the expression of whole regulons under the control of these different σ-factors.

In addition to controlling gene expression through direct interaction with RNAP, the ppGpp has other effects on the cell due to consumption of ATP and GTP for its biosynthesis. ATP is the principal energy carrier in the cell, while GTP is used for polypeptide chain elongation during protein biosynthesis. The depletion of these compounds would lead to immediate arrest of protein biosynthesis as well as most of the other growth-associated metabolic processes. In *Streptomyces coelicolor* A3(2) a five-fold decrease in the intracellular concentration of GTP was observed 7.5 min after nutritional shift-down, while the level of ATP remained essentially unchanged [Chakraburtty et al, 1997]. The Δ*relA* mutant responded to the shift-down by increase in GTP concentration. The absence of effect on ATP level is surprising, but the authors continued observing the nucleotides levels only 15 min after the nutritional shift-down and because some carry-over of the nutrients is highly likely, the measurements perhaps should have been made for a longer time before conclusions on ATP concentration changes can be made. In another work, where the stringent response was induced by serine hydroxamate, more profound changes in nucleotide concentrations have been observed: 20 min after SHX addition GTP and UTP levels decreased to 20-25% and ATP and CTP to 50% of the initial values [Strauch et al, 1991].

6.3 Materials and methods

6.3.1 Chemicals

All chemicals were of chemical purity degree or higher, purchased from Merck (NJ, USA) or Sigma-Aldrich Co. (MO, USA). Casamino acids were from Difco (Becton,

Dickenson and Company, USA). (The $^{13}C_6$-labeled glucose was from Omicron Biochemicals Inc. (IN, USA).

6.3.2 Strain and culture conditions

We used *Streptomyces coelicolor* A3(2) plasmidless strain M600 ($SCP1^-$, $SCP2^-$) and its congenic Δ*relA* mutant M570 [Chakraburtty et al, 1997]. The spores were obtained on soya flour-mannitol agar plates as described in [Kieser et al, 2000] and stored in 20% glycerol. An equivalent of $1.2 \cdot 10^{11}$ spores were pregerminated for 8 hours at 150 rpm in 50 ml of 2xYT in 500 ml shake flasks with baffles and with added 30 glass balls of Ø3 mm. The germinated spores were harvested by centrifugation for 5 min at 4,000xg, washed with fermentation medium and used for inoculation of 1L-working volume fermentor (Applicon Dependable Instruments, Holland). The medium contained: 10 mM NaH_2PO_4, 20 mM NH_4Cl, 10 mM KCl, 2 mM Na_2SO_4, 2 mM citric acid as chelating agent, 1.25 mM $MgCl_2$, 0.05 mM $CaCl_2$, as well as per liter: 15 g glucose, 5 ml trace elements solution (20 mM $FeCl_3$, 1 mM $CuCl_2$, 5 mM $ZnCl_2$, 10 mM $MnCl_2$, 0.02 mM Na_2MoO_4, 2 mM $CoCl_2$, 1 mM H_3BO_4), 2 g Difco casamino acids (Becton, Dickenson and Company, USA) and 100 µl organic antifoam 204 (Sigma-Aldrich, USA). For labeled substrate cultivations 20% of glucose was fully labeled with C13 isotope – $^{13}C_6$-glucose.

Fermentors were autoclaved with water and antifoam at 121ºC for 40 min and all other components were subsequently added by sterile filtration in the form of stock solutions. The aeration rate was 1 volume per volume per minute. For carbon labeling experiments the inlet air was bubbled through a 2 N NaOH solution in order to remove atmospheric carbon dioxide, however this was done only after the cultures started growing and the carbon dioxide percentage in the outlet air was about 0.06%. The pH of the medium was kept at 6.8-7.0 by automatic addition of 0.1 N NaOH and the temperature was kept at 30°C throughout the cultivation.

6.3.3 Biomass dry weight

8-10 ml of culture broth was filtered through a 0.45-µm pore size pre-dried filter (Supor-450, Pall Corporation, Ann Arbor, USA). The filtrate was collected and used for metabolites analysis on HPLC and for analysis of antibiotics. The biomass on the filter was washed twice with distilled water; the filter with biomass was then dried in a microwave oven at 170 W for 20 min, cooled down in a desiccator for 15 min and the mass gain was measured.

6.3.4 Analysis of glucose, amino acids and extracellular metabolites

Glucose and several extracellular primary metabolites were analyzed by HPLC with an Aminex HPX-87H column (Bio-Rad Laboratories, Hercules, CA, USA) operating at 60°C. The compounds were eluted with 5 mM H_2SO_4 at a flow rate of 0.6 $ml \cdot min^{-1}$. Glucose, ethanol, glycerol and succinate were quantified using a differential refractometer (Waters 410; Millipore, Bedford, MA, USA) whereas acetate and pyruvate were quantified using a tunable absorbance detector set at 210 nm (Waters 486; Millipore, Bedford, MA, USA).

For analysis of amino acids, a filtered sample was concentrated by freeze-drying and injected into HPLC. After separation on ion-exchange column, the amino acids were oxidized, derivatized with orto-phthalaldehyde and detected by amperometric detector. The procedure is described in details in [Barkholt et al, 1989].

For analysis of the blue actinorhodin pigment we added 2 ml of 3N KOH to 2 ml of whole broth and extracted with shaking at 4℃ overnight. Then we centrifuged the mixture at 5,000xg for 15 min and measured the absorbance of supernatant at 640 nm. The concentration was calculated using extinction coefficient of 25,320 $M^{-1}cm^{-1}$. The extraction of red prodigiosin pigments was performed according to the same procedure except that we used 2 ml of methanol buffered with HCl to pH 1.5 for extraction. The absorbance was measured at 530 nm and extinction coefficient was 100,500 $M^{-1}cm^{-1}$.

6.3.5 ppGpp analysis

For analysis of ppGpp 50 ml of sample was quickly withdrawn from fermentor, added to the tube with a bit of crashed ice, mixed and centrifuged for 4,000xg for 1 min. The supernatant was decanted and the pellet was frozen in liquid nitrogen and kept at -80℃ until further analysis. 4 ml of 1M formic acid was added to the pellet, the sample was vortexed and sonicated in ice-cold water bath for 10 min. The extraction was continued for 2 hours on a rotating mixer in a cold room (4℃). The sample was filtered through nitrocellulose membrane with pore size 0.45 μm, the filtrate was collected, frozen in liquid nitrogen and freeze-dried. The dry matter was dissolved in 200 μl of MilliQ water and stored at -20℃. The quantification of ppGpp was carried out by HPLC with Partisil 10-SAX column (Whatman, UK) as described in [Takahashi et al, 2004].

6.3.6 GC-MS analysis

From the labeled substrate cultivations samples of 10-20 ml were taken for the analysis of cell mass labeling. Samples were centrifuged at 4000 rpm for 2 min, washed twice with 0.9% NaCl and stored at -20℃ until further analysis.
15 mg of wet cell mass was hydrolyzed with 6M HCl at 105 ºC for 24 hours to release free amino acids from protein molecules. The hydrolysate was centrifuged at 15,000xg to remove the cell debris, divided into two aliquots and dried at 105℃. The crude hydrolysates were subjected to derivatization that rendered them volatile: amino acids were converted to N-ethoxycarbonyl amino acid ethyl esters and N-(N,N-dimethylaminomethylene) amino acid methyl esters and glucose to glucose pentaacetate [Christensen et al, 1999]. Analysis was made by gas chromatography coupled with mass spectrometry as previously described [Christensen et al, 1999]. The signal intensities were corrected for occurrence of natural isotopes in the atoms of the derivative part and in oxygen and nitrogen atoms of the amino acids. The corrected intensities were used for calculating the summed fractional labeling (SFL) of a fragment according to the formula: $SFL = \frac{i_0 \cdot 0 + i_1 \cdot 1 + i_2 \cdot 2 + \dots i_j \cdot m_j \dots + i_n \cdot n}{i_0 + i_1 + i_2 + \dots i_j \dots + i_n} \times 100$, where *ij* is the peak intensity of mass isotopomer *mj* and *n* is the number of carbon atoms in the fragment. The SFL value shows the amount of ^{13}C atoms per 100 fragments.
In order to compare the labeling states of two metabolites with different numbers of carbon atoms carbon normalized labeling (CNL) was calculated as: $CNL = \frac{SFL}{n}$, where SFL is the summed fractional labeling of all the carbons in the fragment and *n* is the number of carbon atoms in the fragment.

6.3.7 RNA isolation and gene expression analysis

The samples of 20 ml were withdrawn quickly and sprayed into 50 ml falcon tubes filled with crushed ice resulting in almost immediate cooling down to 0-2 ºC. The mixture was shaken and centrifuged at 4,000xg for 1 min. The supernatant was decanted, the cells were immersed in liquid nitrogen and then stored at -80ºC until RNA extraction. The total time spent on the sampling procedure did not exceed two min.

Total RNA extraction was performed with RNeasy kit (QIAGEN GmbH, Germany) using 3x40s cycles on FastPrep FP120 (Qbiogene Inc., USA) for breaking the cells. After separation of cellular debris, additional steps were added: two rounds of phenol:chlorophorm:isoamylalcohol (pH8) extraction and 1 round of chlorophorm extraction. The rest of the procedure was according to the manufacture's instructions. The total RNA was subjected to DNase digestion and then concentrated from the reaction solution using RNeasy CleanUp kit (QIAGEN GmbH, Germany). The quality of RNA was assessed on Bioanalyzer (Agilent, USA) to make sure there was no contamination with DNA and no degradation of RNA occurred. The RNA samples were labeled and hybridized to the arrays according to the procedure described on http://www.surrey.ac.uk/SBMS/Fgenomics/Microarrays. The arrays were spotted oligonucleotide arrays, containing duplicate spots of probes for 7337 *S. coelicolor* genes. Each probe was 50 nucleotide bases long and was designed for minimal cross-hybridization with other genes. To maximize the usage of arrays we applied look design [Vinciotti et al, 2005] with using a genomic DNA as a common reference (Appendix 6.1). The arrays were scanned with Affymetrix 428 laser scanner (Affymetrix Inc., USA). Image analyses was made using BlueFuse software (BlueGnome Limited, UK), which allows automatic grid alignment, spots boundaries detection, background substraction and signal intensities estimation. The spots quality was scored to 0 or 1 depending on their radius, circularity, uniformity and grid offset. For normalization and data analysis we used maanova package [Cui et al, 2003] implemented in R statistical programming environment (http://www.r-project.org). Maanova stands for microarray analysis of variance and is well suited for experimental designs with multiple factors. The intensities for duplicate spots were averaged, log transformed and normalized within array using glowess [Kerr et al, 2000]. Further on the intensities were fit an ANOVA model according to [Cui et al, 2003]. ANOVA model is fit in two stages, the first stage is the normalization model:

$$y_{ijkgr} = \mu + A_i + D_j + AD_{ij} + r_{ijkgr} \quad (1)$$

The term μ captures the overall mean. The other terms capture the overall effect of array (A), dyes (D) and labeling reactions (AD). The residual of the first stage is then used as the input for the second stage. The second stage models gene-specific effects:

$$r_{ijkr} = G + AG_i + DG_j + VG_k + \varepsilon_{ijkr} \quad (2)$$

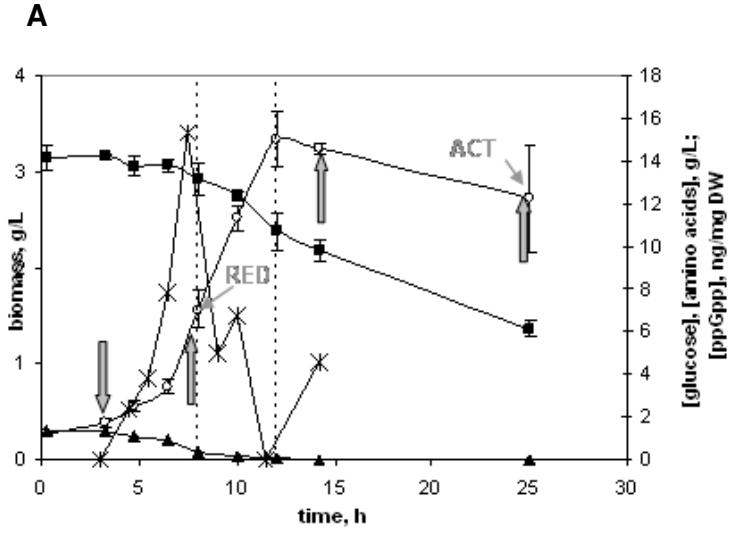
Here G captures the average effect of the gene. AG captures the array by gene variation. DG captures the dye by gene variation. VG captures the effects for the experimental varieties. To compare two conditions we used permutation test [Kerr et al, 2000], also implemented in maanova. Consensus clustering was done using a macluster function in maanova package with 1000 permutations at 0.6 significance.

6.4 Results

6.4.1 ppGpp slows down the growth rate and leads to more efficient nutrient utilization

S. coelicolor strain M600 and its congenic Δ*relA* mutant were grown in well-controlled fermentors in mineral medium supplemented with casamino acids. Because ppGpp response is best induced by starvation for amino acids, we used growth-limiting concentrations of nitrogen, half of nitrogen was inorganic ammonium chloride and half was coming from casamino acids. Ammonium and amino acids were consumed by the cells concurrently, though amino acids were exhausted before ammonium (data not shown). No overflow metabolism, signified by secretion of primary metabolites as pyruvate, acetate or similar, was observed in any cultivation.
The strain with deleted *relA* gene grew faster than the reference strain both in the first and the second exponential growth phases (Figure 6-1, Table 6-1). It also consumed more glucose and amino acids for synthesis of a unit of biomass, which resulted in a lower biomass concentration at the end of the growth phase in the mutant cultivation in comparison to the reference.
The intracellular concentration of ppGpp reached maximum around the end of the 1st exponential growth phase of M600 and that was also the time when pigmented prodigiosins could be detected in the medium (Figure 9-1). Actinorhodin pigments were first detected in the stationary phase. As expected, the M570 did not produce ppGpp at any time of cultivation. The onset of RED production in M570 was delayed, it appeared first in the stationary phase. The transcription of RED biosynthetic cluster activator redD reached the maximum at the 2nd point in M600 and only at the 3rd point in M570.

Figure 6-1. Fermentation profiles of *S. coelicolor* reference (A) and *ΔrelA* (B) in batch cultivations in N-limited minimal medium supplemented with casamino acids, glucose - ■, biomass dry weight - ○, total amino acids - ▲, ppGpp - ж. The arrows show time points when samples for mRNA were taken. Error bars show standard deviations calculated from biological triplicates. Times of onset of undecylprodigiosin and actinorhodin production are indicated by RED and ACT signs correspondingly.



B

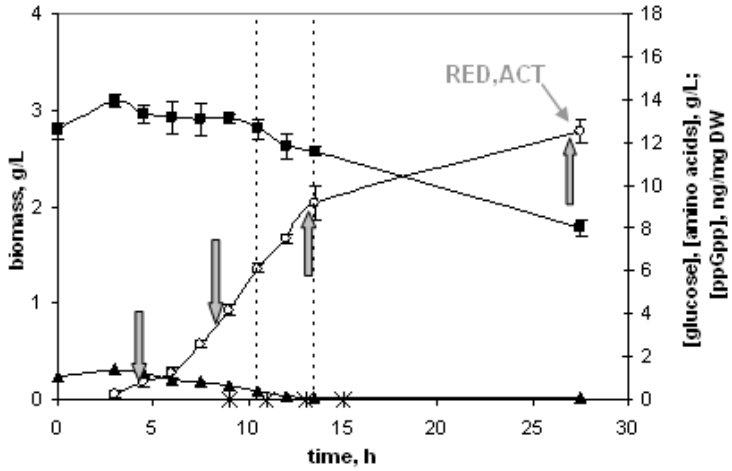


Table 6-1. Physiological parameters of *S. coelicolor* reference strain (M600) and its Δ*relA* mutant (M570)

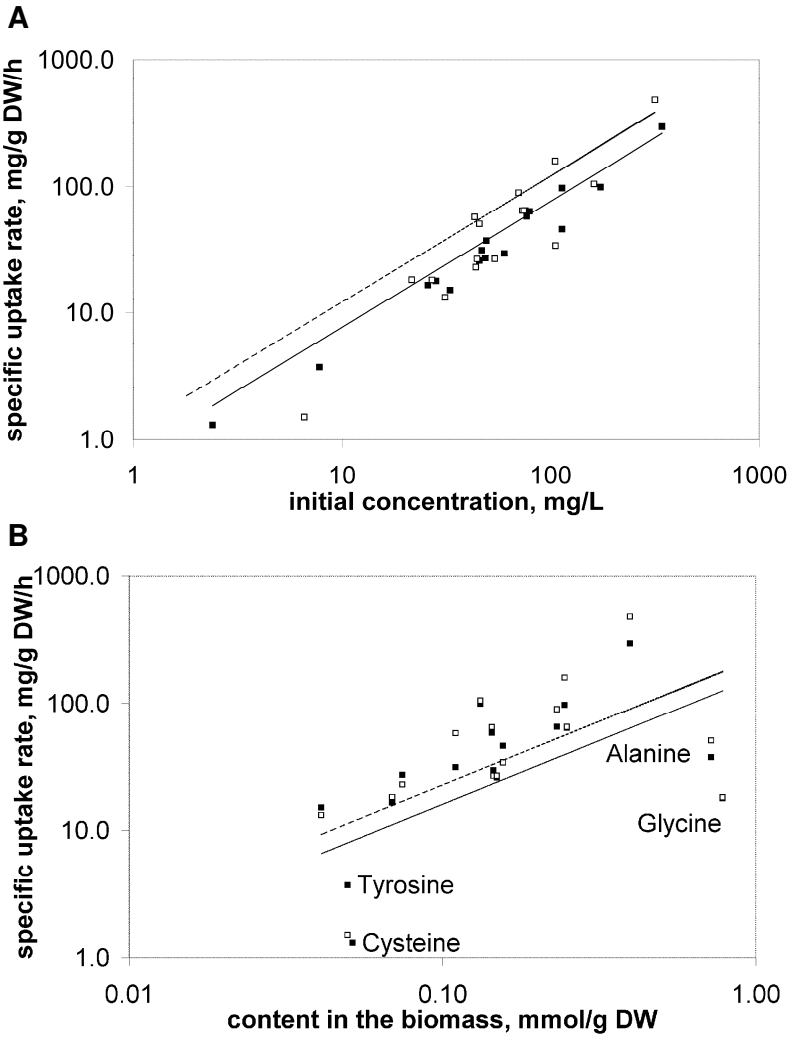
Parameters[c]	M600		M570	
	Phase 1[a]	Phase 2[b]	Phase 1[a]	Phase 2[b]
Maximal specific growth rate, μ_{max}, h^{-1}	0.30±0.02	0.19±0.04	0.35±0.03	0.13±0.02
Specific glucose uptake, Y_{xs}, (g glucose)·(g dry weight)	0.83	1.50±0.29	1.32	1.58±0.31
Specific amino acids uptake, Y_{xaa}, (g amino acids)·(g dry weight)	0.93	0.14	1.24	0.35

[a] Exponential growth on amino acids, ammonium and glucose
[b] Exponential growth on ammonium and glucose
[c] The numbers are average values between four biological replicates (three non-labeled and one labeled cultivation), standard deviations are also shown. Where standard deviation is not shown, the numbers were only measured in labeled cultivations.

6.4.2 Amino acids uptake and biosynthesis are weakly regulated

All amino acids were uptaken simultaneously; there was no multiphasal growth on preferred amino acids. The specific consumption of amino acids correlated linearly with their initial concentration in the cultivation medium (Figure 6-2A), pointing to the presence of amino acids-nonspecific transport complexes.

Figure 6-2. Correlation between specific uptake rate of amino acids and their initial concentration in the medium (A). Correlation between specific uptake rate of amino acids and amino acids content in the biomass (B). ■ – M600 data, □ – M570 data. The data about content in the biomass was calculated based on protein composition [Davidson, 1992] and protein and peptidoglycan content [Borodina et al, 2005] at dilution rate 0.3h^{-1}.



The specific uptake rates of amino acids were also proportional to the content of a specific amino acid in the biomass, with the exception of alanine and glycine (Figure 6-2B). Alanine and glycine are used in the biosynthesis of peptidoglycan, which explains their high abundance in *Streptomyces* biomass.

The specific consumption of most of the amino acids exceeded the content of amino acids per unit of biomass several fold (Table 6-2). This shows that the uptaken amino acids were used not only as building blocks for protein biosynthesis but also as a source of energy and precursors, because larger part of them was degraded. Consumption of aspartate, glutamate and arginine was 50-60% higher in the mutant than in the reference strain.

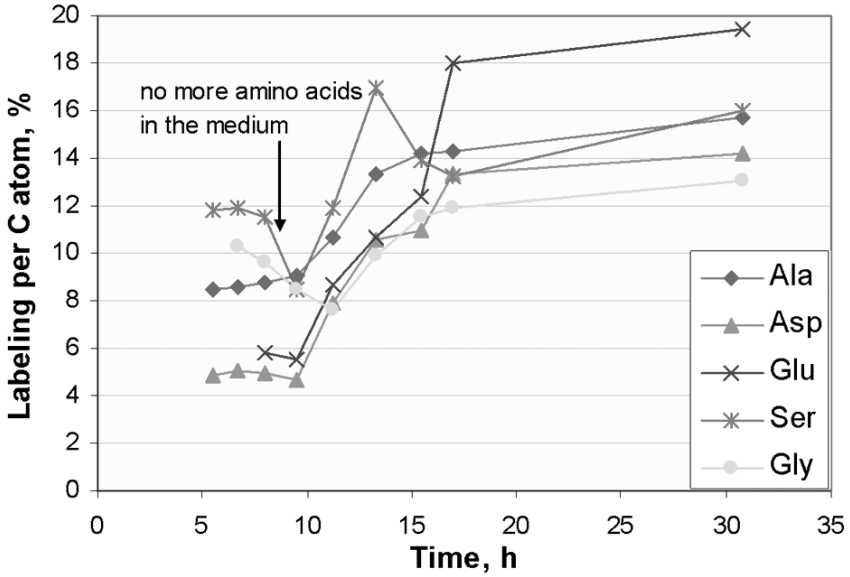
Table 6-2. Content of amino acids in the biomass and their specific uptake.

Amino acid	Content in the biomass, mmol/g DW	Specific uptake, mmol/g DW	
		M600	M570
Aspartate/Asparagine	0.25	2.87	3.99
Threonine	0.15	0.99	0.76
Serine	0.14	2.27	2.13
Glutamate/Glutamine	0.40	7.81	10.79
Proline	0.13	3.46	3.11

Glycine	0.78	1.06	0.92
Alanine	0.72	1.78	2.05
Cysteine	0.05	0.04	0.00
Valine	0.25	2.18	1.88
Metionine	0.07	0.43	0.40
Isoleucine	0.15	0.78	0.68
Leucine	0.23	1.94	2.27
Tyrosine	0.05	0.08	0.03
Phenylalanine	0.07	0.63	0.45
Histidine	0.04	0.37	0.28
Lysine	0.16	1.22	0.77
Arginine	0.11	0.68	1.06

To further study amino acids consumption we analyzed the labeling of proteinogenic amino acids from the cultivation on 20% $_6C^{13}$-labelled glucose. We took several biomass samples in the course of cultivation, both in the exponential and stationary phases, hydrolyzed the proteins and analyzed the labeling of the proteinogenic amino acids on GC-MS.
The labeling of amino acids was constant in the exponential phase of cultivation (Figure 6-3); after exhaustion of amino acids in the medium the labeling began to increase indicating biosynthesis from labeled glucose and then the labeling almost reached a new plato.

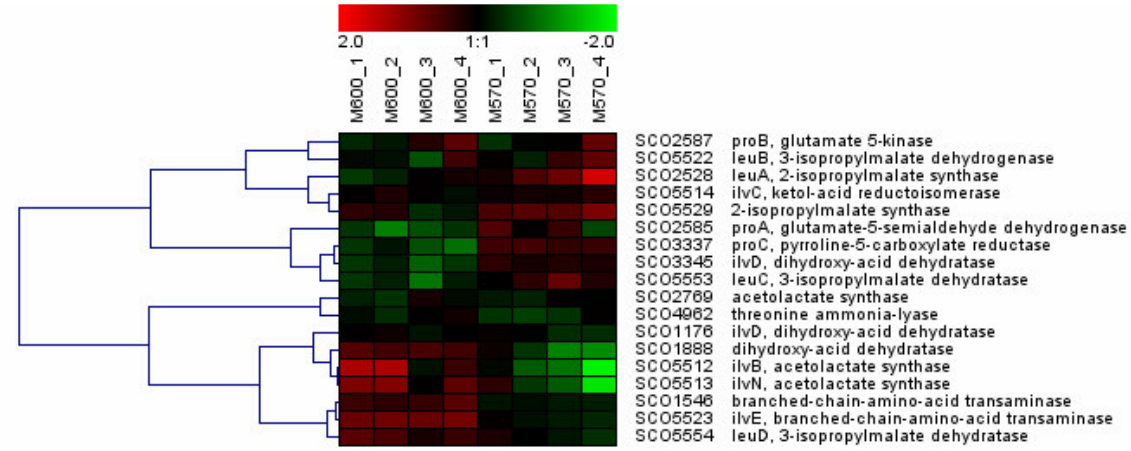
Figure 6-3. Dynamics of label incorporation into proteinogenic amino acids (the patterns for both strains were similar, here strain 570 is shown).



A few amino acids (leucine, isoleucine, proline, valine) were only naturally labeled (Table 6-3), meaning that these amino acids were not synthesized when there was sufficient supply from the medium. The first reaction in proline biosynthesis is catalyzed by glutamate-5-kinase, which is encoded by gene *proB* (SCO2587). This gene is significantly downregulated in the early time points of cultivation both in the reference and in the mutant strain (Figure 6-4). Under different experimental conditions it was found that proline genes are constitutively expressed and are

stimulated during growth on proline [Hood et al, 1992], hence their regulation is conditional.
The precursor for biosynthesis of branched chain amino acids valine and leucine is 2-acetolactate and for isoleucine it is 2-aceto-2-hydroxy-butanoate, both of these precursors are made by the action of acetolactate synthase. The latter consists of small and large subunits *ilvN* (SCO5513) and *ilvB* (SCO5512). Open reading frame SCO2769 was annotated as putative acetolactate synthase. While the expression of SCO2769 increases in the time course of cultivation, the expression of *ilvB* and *ilvN* decreases and is higher in the reference than in the mutant strain. The expression of these two genes correlates very well, they make an operon, which is transcribed from promoter upstream *ilvB*. The expression from *ilvB* promoter was shown to decrease in the presence of the three branched amino acids [Craster et al, 1999]. From our results we can conclude that the expression of *ilvBN* decreased at the appearance of ppGpp, but there was no repression of genes by casamino acids. The synthesis of branched amino acids could have been inhibited on the enzymatic level: it's known that the small subunit of acetolactate synthase has a binding site for valine, which inhibits the enzyme [Kopecky et al, 1999; Bauerle et al, 1964].
LeuA and *leuB* genes, both involved in leucine synthesis were dowregulated during the growth on amino acids. So was threonine ammonium lyase (SCO4962), which is needed for biosynthesis of 2-oxobutanoate, isoleucine precursor.

Figure 6-4. Expression of enzymatic genes involved in biosynthesis of proline and branched amino acids.



Interconversions between glyceraldehyde-3-phosphate and glycine, serine and threonine lead to high labeling of these amino acids.
High labeling in alanine, glutamate and aspartate reflect participation of these amino acids in transamination reactions, which equilibrate the labeling in the intracellular pools of amino acids and their precursors: pyruvate, alpha-ketoglutarate and oxaloacetate correspondingly.
Weak labeling of phenylalanine and lysine indicate that the control of their synthesis is not very tight and their presence in the medium does not stop the biosynthesis from the labeled precursors completely. Aspartate kinase, the first enzyme in lysine biosynthesis, has been shown to be under lysine-sensing riboswitch control in *B. subtilis* [Mandal et al, 2003] and in *E. coli* [Patte et al, 1998]. In *B. subtilis* truncated versions of aspartate kinase mRNA are made in the presence of lysine [Mandal et al, 2003].

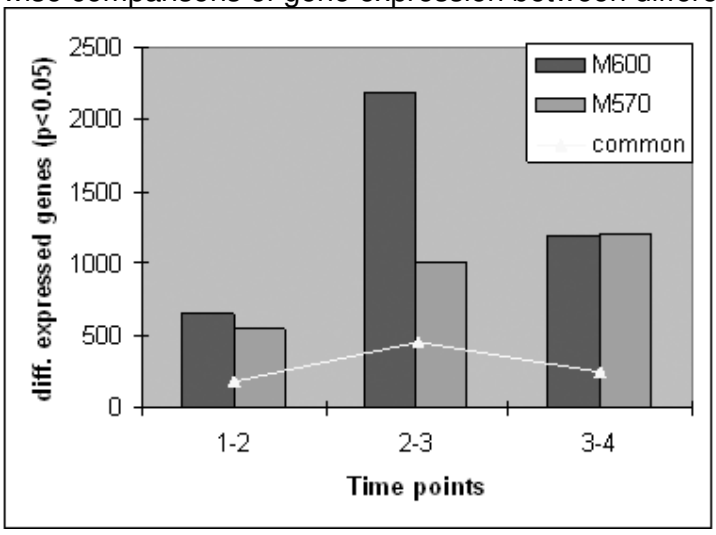
Table 6-3. Labeling of amino acids at the end of the exponential growth on amino acids. (summed fractional labeling in % per carbon atom, average values from several amino acids fragments)

	SFL normalized per atom, %	
Amino acid	M600	M570
Alanine	8.3±0.6	9.3±0.4
Valine	2.3±1.6	2.9±2.1
Leucine	1.1	1.1
Isoleucine	1.5	2.1
Lysine	3.4	3.6
Aspartic acid	6.4±2.3	7.0±2.2
Threonine	3.1±0.6	3.2±0.4
Glycine	9.1±0.6	9.5±0.7
Serine	11.6±3.7	10.3±1.7
Proline	1.1	1.1
Glutamic acid	4.9±1.0	5.7±0.8
Phenylalanine	3.7±0.9	4.7±0.2

6.4.3 Transcriptional changes are largest during transition phase and only in stringent strain

We found how many genes have significantly changed their expression between time points 1 and 2, 2 and 3 and 3 and 4 in the tested strains. The largest changes were observed in the reference strain between time points 2 and 3, which correspond to the transition phase (Figure 6-5). 2200 genes have significantly changed their expression, while in the mutant strain the number was only 900.

Figure 6-5. Pairwise comparisons of gene expression between different time points.



In the context of ppGpp response we were mostly interested in those genes that have changed their expression significantly either between time points 1-2 or 2-3 either in mutant or reference strain or both. We clustered the expression profiles of

those genes using K-means consensus clustering with 1000 iterations (Figure 6-7), the amount of clusters was set to 12 (Figure 6-6).

Figure 6-6. Figure of merit.

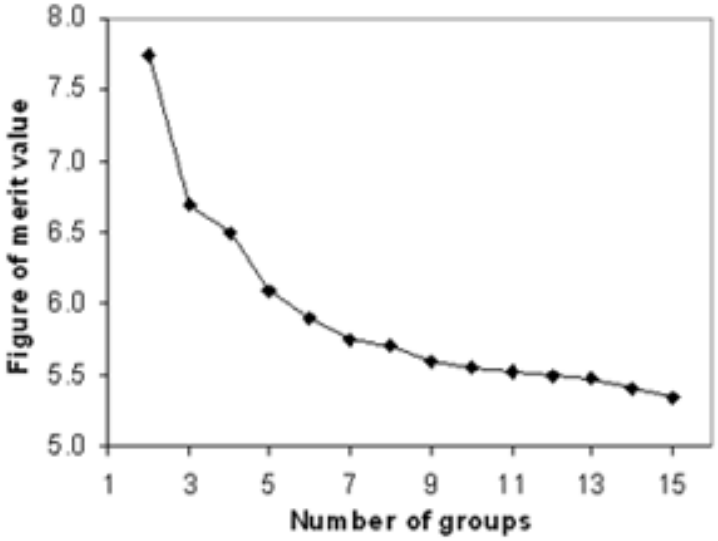
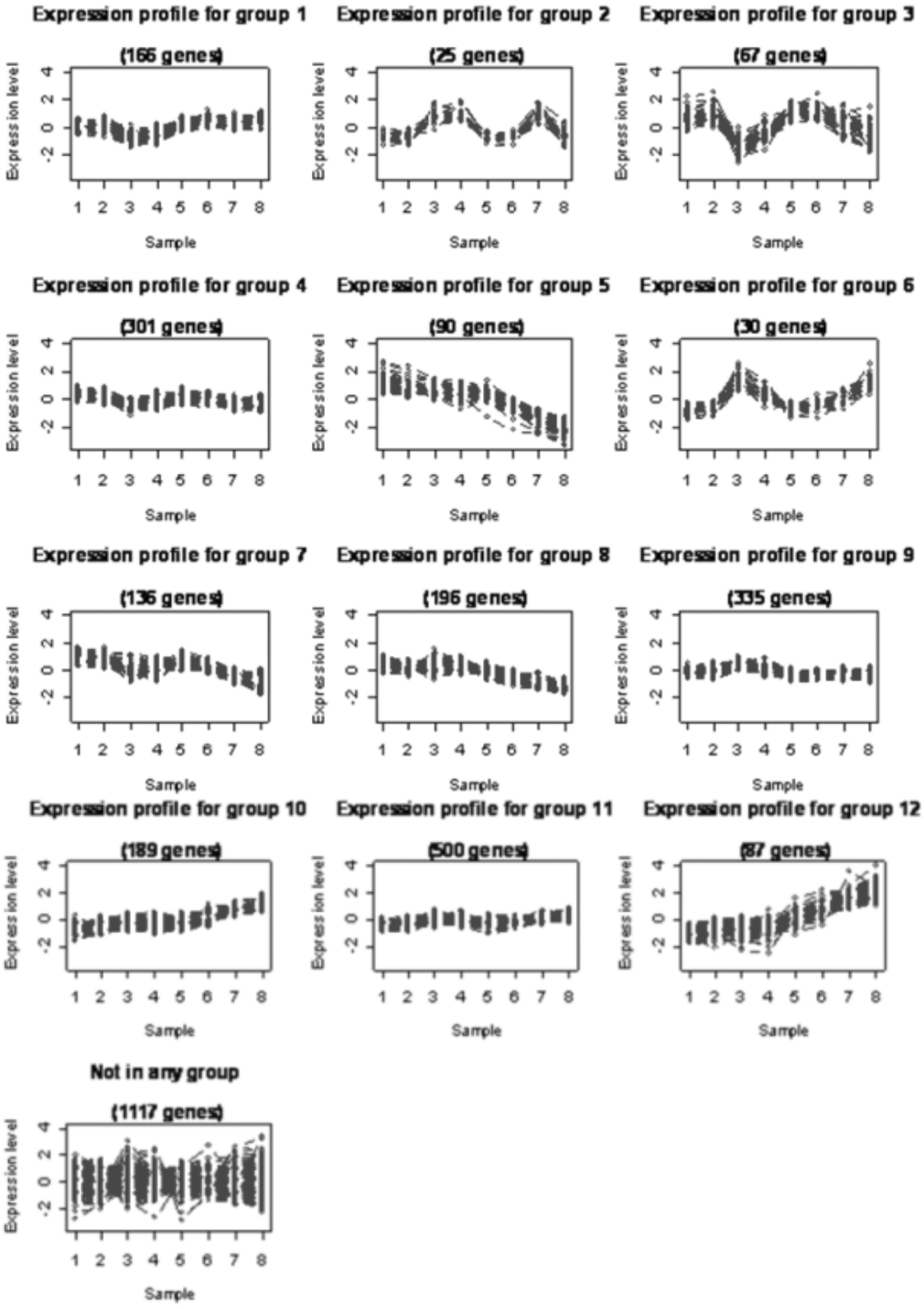


Figure 6-7. Consensus clustering of significantly regulated genes. Samples 1 to 4 represent reference strain M600 samples taken at four time points and samples 5-8 are four time point samples of mutant strain M570.



The genes that have been significantly up- or down–regulated during transition phase in the references strain but not in the mutant strain were sorted into functional categories (up- clusters 6 and 9, down – clusters 1 and 3) (Figure 6-8). The functional categories were as proposed in protein classification scheme by the Sanger Institute (http://www.sanger.ac.uk/Projects/S_coelicolor/scheme.shtml). We hypothesized that these genes directly or indirectly are regulated by ppGpp. The genes with unknown functions are not shown on the graph. The categories with most of the genes are marked by red circles.

Downregulation of genes involved in ribosomes constituents' biosynthesis and modification as well as energy metabolism was observed.

Among the genes involved in energy metabolism, the most represented were cytochrome subunits and ATP synthase chains, while cytochorome P450, known to be involved in secondary metabolism [O'Keefe et al, 1991] was upregulated.

Among the macromolecules synthesis and modification genes the downregulated were tRNA acyl synthases and genes participating in protein translation and modification (peptide chain release factor 1, translation initiation factors IF1 and IF3, elongation factor G).

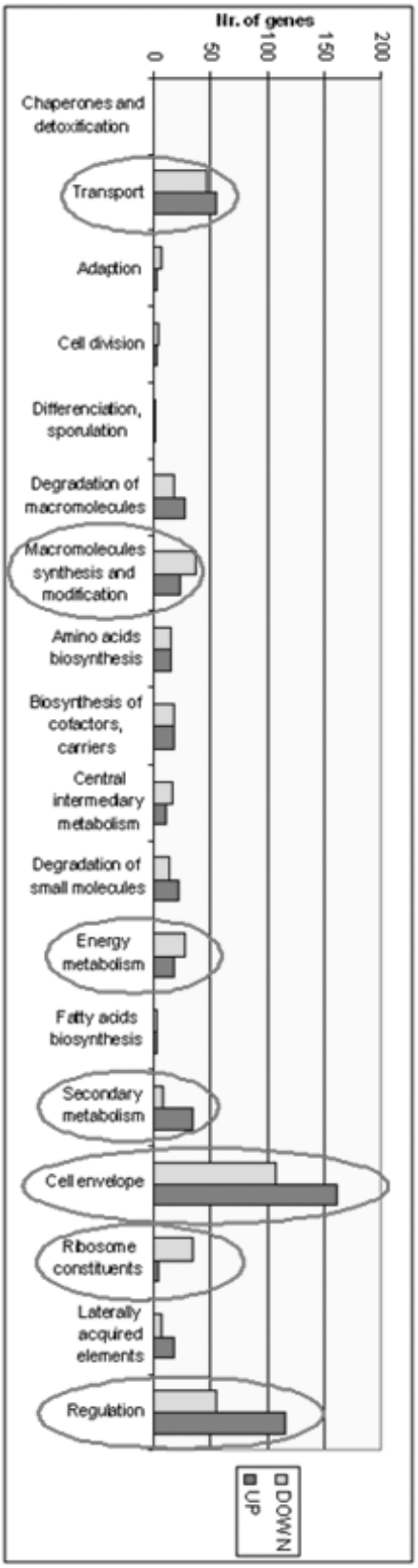
Putative amino acid permease (SCO2914) and glutamate permease (SCO5775) were downregulated, while ammonium transporter (SCO5583) was upregulated, indicating the shift from growth on amino acids to growth on inorganic ammonium.

Secondary metabolism included actinorhodin cluster activator gene, which induces the transcription of actinorhodin cluster [Malpartida et al, 1986; Fernandez-Moreno et al, 1994].

Many changes occurred in genes coding for membrane proteins. Noteworthy even though morphological differentiation with aerial mycelium and spores formation does not occur in liquid cultures, the composition of the membrane still changes.

Multiple regulatory genes were up- or down- regulated in the transition phase in the reference strain, pointing to the fact that ppGpp regulation in *Streptomyces* is a multi-step hierarchical regulation, involving many regulatory proteins. One of the examples is *scbA* gene, which is involved in the synthesis of gamma-butyrolactone signaling molecule SCB1 [Takano et al, 2000; Takano et al, 2001; Butler et al, 2003].

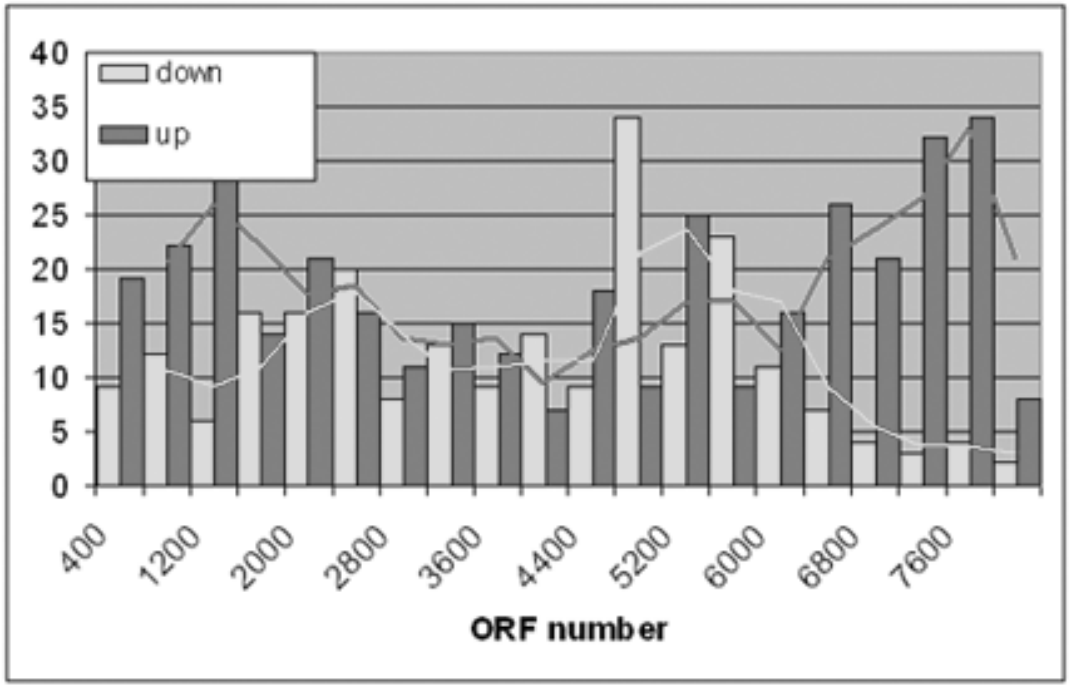
Figure 6-8. Supposedly ppGpp regulated genes sorted by functional categories.



6.4.4 The up- and down- ppGpp-regulated genes are distributed differently on the chromosome

We plotted the ppGpp affected genes according to their position on the chromosome (Figure 6-9). On the arms of the chromosome (left arm 0-1.5 Mb, right arm 6.5-8 Mb), 80% of the genes were upregulated and only 20% were downregulated. On the other hand there was a high concentration of downregulated genes around the origin of replication, positioned at around 4.5 Mb.

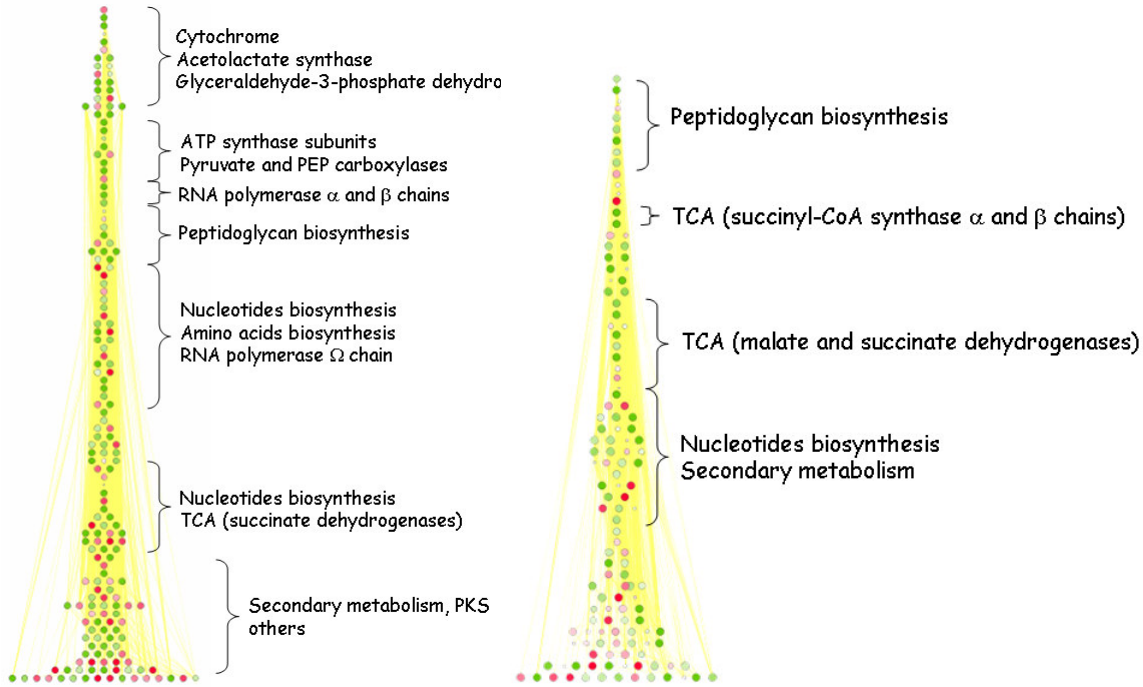
Figure 6-9. Position of the ppGpp-regulated genes on the chromosome.



6.4.6 Most significantly co-regulated metabolic networks

Some genes that are metabolically connected also change their expression together. We identified the significantly co-regulated metabolic subnetworks during transition phase (between time points 2 and 3) in the reference and in the ppGpp-deficient strains (Figure 6-10).

Figure 6-10. Subnetworks of co-regulated metabolic genes (left – reference strain, right – Δ*relA* strain), the networks are displayed in hierarchical layout generated with Cytoscape (www.cytoscape.org). The yellow lines show connections between genes through common metabolites.



We used p-values from t-statistical test comparing time point 2 and 3 within each strain and genome-scale metabolic model reconstructed previously [Borodina et al, 2005]. The significant subnetwork was found using algorithm described by Ideker et. al [Ideker et al, 2002]. The subnetworks for M600 and M570 had almost the same significance score, but the first contained 181 genes and the second 132 genes. The most obvious difference between the networks of the two strains is the absence of cytochrome genes (electron transport chain), ATP synthase subunits and RNA polymerase genes from the network of ppGpp lacking mutant. These genes are crucial for cellular function and apparently ppGpp is important for their coordinated regulation in the transition phase. Otherwise the most significantly regulated genes are the ones from amino acids and nucleotides metabolism as well as peptidoglycan biosynthesis, all the metabolic pathways that involve nitrogen-containing intermediates. Significant regulation of these pathways correlates well with nitrogen exhaustion in the medium in the transitionary phase. It is notable that genes from these pathways are highly regulated both in the reference and mutant strains pointing to ppGpp-independent regulation.

6.5 Discussion

Guanosine 3´-diphosphate, 5´-diphosphate is a global signaling molecule, which was originally found in bacteria but recently also observed in the chloroplasts of higher plants [Takahashi et al, 2004]. The intracellular concentration of ppGpp is regulated by ppGpp synthase and hydrolase and can increase very rapidly in response to stress. Some residual amounts of ppGpp are always present in the cells. We observed that the *S. coelicolor* strain, which had a deletion of ppGpp synthase and didn't have any residual amounts of ppGpp at any stage of cultivation, was growing slightly faster than the reference strain but was consuming more glucose and amino acids to generate a unit of dry weight biomass. In contrast, *E. coli* with deleted *relA* and *spoT* genes was shown to grow 10-30% slower than the wild type strain on several complex media [Hernandez et al, 1993], it was not, however, investigated if

the specific biomass yield was changed. On the other hand, overexpression of *relA* in *E. coli* caused inhibition of rRNA synthesis and slower growth [Schreiber et al, 1991; Tedin et al, 1992].
Our labeling experiments showed that biosynthesis of many amino acids occurred even while there was sufficient supply from the medium, meaning weak regulation on transcriptional/enzymatic levels. Exception was biosynthesis of proline and branched chain amino acids, which were not synthesized *de novo*, while the cells grew on amino acids. As the labeling of amino acids was very similar in the reference and mutant strains, we can conclude that ppGpp does not participate in regulation of amino acids biosynthesis in *S. coelicolor*. Interestingly, in *E. coli* ppGpp is essential for cells in order to synthesize amino acids, therefore the ppGpp-deficient *E. coli* mutants are auxotrophic for several amino acids [Xiao et al, 1991; Imaizumi et al, 2006]. Weakly regulated amino acids biosynthesis can also explain why *S. coelicolor* Δ*relA* mutants do not have a growth defect in contrast to *E. coli*.
When *S. coelicolor* goes from exponential into stationary growth phase and intracellular ppGpp concentration peaks, more than one fifth of the genes change their expression level significantly. Most of these changes were not observed in the $ppGpp^-$ strain. The functional categories that are particularly effected are involved in stable RNA biosynthesis and secondary metabolism, which has been described before [Strauch et al, 1991; Chakraburtty et al, 1997]. What was not known before was an extensive effect on the cellular envelope constituents and on regulatory proteins. During growth on complex medium morphological differentiation occurs on nutrients starvation, therefore the composition of the cellular envelope is expected to change. In submerged cultures, however, *S. coelicolor* normally doesn't sporulate and no particular changes in morphology can be detected microscopically. *S. coelicolor* has over a thousand genes that are putative transcriptional regulators, obviously part of the ppGpp effect is generated through these regulators, meaning that ppGpp, besides direct regulation of transcription, starts a signaling cascade, which results in the stringent response. Some groups of metabolically connected genes were found to be significantly regulated in the reference strain, but not in the mutant. These were cytochrome-coding genes involved in electron transport chain, ATP-synthase subunits and RNA polymerase genes.
Interestingly, at the arms of the chromosomes the genes up-regulated by ppGpp were more abundant, with 80% of the genes being up-regulated. There was a high concentration of down-regulated genes around the origin of replication. It's been shown previously by Karoonuthaisiri et al. that the core of the chromosome is downregulated in the stationary phase, but the authors did not find a trend in upregulation of chromosome arms, they rather saw genes across all the genome being up-regulated [Karoonuthaisiri et al, 2005].

6.6 Conclusions

Antibiotics production in *Streptomyces* is under a pleotrophic regulation of many transcription factors, σ-factors, signaling molecules and global regulatory proteins. ppGpp is an important global regulator as it most probably influences the binding of numerous σ-factors to promoters, therefore it is an interesting object for future studies.
The cascade of regulatory events is so complicated in *Streptomyces* that it'll take many more years before we'll be able to decipher it even for the most studied model streptomycete – *S. coelicolor* A3(2). As reconstruction of regulatory pathways is

simply non-feasible for all the industrially interesting *Streptomyces* species, one might hope that some general rules will be discovered in *S. coelicolor* A3(2) that will help us improve antibiotics production in other strains. So far the work on the regulation in *Streptomyces* mainly concentrated on separate major regulators or regulatory cascades. Now the genomic techniques that have become available for *S. coelicolor* open up possibilities for a more structured and systematic approach. We hope that directed efforts from a number of scientific groups interested in *Streptomyces* will be coordinated and will result in a deeper understanding of regulation in these complex bacteria.

6.7 Acknowledgements

We thank Dr. Ochi for presenting us with the ppGpp standard, Hanne Jacobsen for assistance with setting up HPLC system for ppGpp analysis, Anne Blicher for performing amino acids quantification and Technical University of Denmark for financial support.

6.8 References

Aigle B, Wietzorrek A, Takano E, Bibb MJ. (2000). "*A single amino acid substitution in region 1.2 of the principal sigma factor of Streptomyces coelicolor A3(2) results in pleiotropic loss of antibiotic production*." *Mol Microbiol*, **37**(5):995-1004.

An G, Vining LC. (1978). "*Intracellular levels of guanosine 5'-diphosphate 3'-diphosphate (ppGpp) and guanosine 5'-triphosphate 3'-diphosphate (pppGpp) in cultures of Streptomyces griseus producing streptomycin*." *Can J Microbiol*, **24**(5):502-511.

Barkholt V, Jensen AL. (1989). *"Acid analysis: determination of cysteine plus half-cystine in proteins after hydrochloric acid hydrolysis with a disulfide compound as additive." Analytical Biochem*, **177**:318-322.

Bauerle RH, Fruendlich M, Stormer FC, Umbarger HE. (1964). "*Control of isoleucine, valine and leucine biosynthesis. II. Endproduct inhibition by valine of acetohydroxy acid synthetase in Salmonella typhimurium*." *Biochim Biophys Acta*, **92**:142-149.

Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D *et al*. (2002). "*Complete genome sequence of the model actinomycete Streptomyces coelicolor* A3(2)." *Nature*, **417**(6885):141-147.

Borodina I, Krabben P, Nielsen J. (2005). "*Genome-scale analysis of Streptomyces coelicolor* A3(2) metabolism." *Genome Res*, **15**(6):820-829.

Butler MJ, Takano E, Bruheim P, Jovetic S, Marinelli F, Bibb MJ. (2003). "*Deletion of scbA enhances antibiotic production in Streptomyces lividans*." *Appl Microbiol Biotechnol*, **61**(5-6):512-516.

Cashel M, Gallant J. (1969). "*Two compounds implicated in the function of the RC gene of Escherichia coli*." *Nature*, **221**(5183):838-841.

Cashel M, Hsu LM, Hernandez VJ. (2003). "*Changes in conserved region 3 of Escherichia coli sigma 70 reduce abortive transcription and enhance promoter escape*." *J Biol Chem*, **278**(8):5539-5547.

Chakraburtty R, White J, Takano E, Bibb M. (1996). "*Cloning, characterization and disruption of a (p)ppGpp synthetase gene (relA*) of *Streptomyces coelicolor* A3(2)." *Mol Microbiol*, **19**(2):357-368.

Chakraburtty R, Bibb M. (1997). "*The ppGpp synthetase gene (relA*) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation." *J Bacteriol*, **179**(18):5854-5861.

Chatterji D, Fujita N, Ishihama A. (1998). "*The mediator for stringent control, ppGpp, binds to the beta-subunit of Escherichia coli RNA polymerase*." *Genes Cells*, **3**(5):279-287.

Christensen B, Nielsen J. (1999). *"Isotopomer analysis using GC-MS." Metab Eng* **1**(4):282-290.

Craster HL, Potter CA, Baumberg S. (1999). "*End-product control of expression of branched-chain amino acid biosynthesis genes in Streptomyces coelicolor A3(2): paradoxical relationships between DNA sequence and regulatory phenotype*." *Microbiology*, **145**:2375-2384.

Cui X, Churchill GA. (2003). "*Statistical tests for differential expression in cDNA microarray experiments*." *Genome Biol*, **4**(4):210.

Davidson, A. (1992). *"Quantitative Microbial Physiology of Streptomyces coelicolor A3(2)"*, PhD thesis, University of Glasgow, Glasgow.

Fernandez-Moreno MA, Martinez E, Caballero JL, Ichinose K, Hopwood DA, Malpartida F. (1994). "*DNA sequence and functions of the actVI region of the actinorhodin biosynthetic gene cluster of Streptomyces coelicolor A3(2).*" *J Biol Chem*, **269**(40):24854-24863.**Haseltine WA, Block R**. (1973). "*Synthesis of guanosine tetra- and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes*." *Proc Natl Acad Sci U S A* , **70**(5):1564-1568.

Hernandez VJ, Bremer H. (1993). "*Characterization of RNA and DNA synthesis in Escherichia coli strains devoid of ppGpp*." *J Biol Chem*, **268**(15):10851-10862.

Hernandez VJ, Cashel M. (1995). "*Changes in conserved region 3 of Escherichia coli sigma 70 mediate ppGpp-dependent functions in vivo*." *J Mol Biol*, **252**(5):536-549.

Hesketh A, Sun J, Bibb M. (2001). "*Induction of ppGpp synthesis in Streptomyces coelicolor A3(2) grown under conditions of nutritional sufficiency elicits actII-ORF4 transcription and actinorhodin biosynthesis*." *Mol Microbiol*, **39**(1):136-144.

Hood DW, Heidstra R, Swoboda UK, Hodgson DA. (1992). "*Molecular genetic analysis of proline and tryptophan biosynthesis in Streptomyces coelicolor A3(2): interaction between primary and secondary metabolism--a review*." *Gene*, **115**(1-2):5-12.

Ideker T, Ozier O, Schwikowski B, Siegel AF. (2002). "*Discovering regulatory and signalling circuits in molecular interaction networks.*" *Bioinformatics*, **18**:233-240.

Imaizumi A, Kojima H, Matsui K. (2006). "*The effect of intracellular ppGpp levels on glutamate and lysine overproduction in Escherichia coli*." *J Biotechnol*, **125**(3):328-337.

Jishage M, Kvint K, Shingler V, Nystrom T. (2002). "*Regulation of sigma factor competition by the alarmone ppGpp*." *Genes Dev*, **16**(10):1260-1270.

Karoonuthaisiri N, Weaver D, Huang J, Cohen SN, Kao CM. (2005). "*Regional organization of gene expression in Streptomyces coelicolor*." *Gene*, **353**(1):53-66.

Kerr MK, Martin M, Churchill GA. (2000). *"Analysis of variance for gene expression microarray data". J Comp Biol*, **7**(6):819-837.

Kieser T, Bibb M, Buttner M, Chater K, D.A. H. (2000). "*Practical Streptomyces Genetics*." The John Innes Foundation

Kopecky J, Janata J, Pospisil S, Felsberg J, Spizek J. (1999). "*Mutations in two distinct regions of acetolactate synthase regulatory subunit from Streptomyces cinnamonensis result in the lack of sensitivity to end-product inhibition*." *Biochem Biophys Res Commun*, **266**(1):162-166.

Malpartida F, Hopwood DA. (1986). "*Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in Streptomyces coelicolor A3(2)*." *Mol Gen Genet*, **205**(1):66-73.

Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. (2003). "*Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria*." *Cell*, **113**(5):577-586.

Muta S, Osoegawa K, Ezaki S, Zubair M, Kuhara S, Mukai J, Dixon R. (1992). "*Streptomyces ATP nucleotide 3'-pyrophosphokinase and its gene*." *Nucleic Acids Symp Ser,* (27):165-166.

O'Keefe DP, Harder PA. (1991). *"Occurrence and biological function of cytochrome P450 monooxygenases in the Actinomycetes." Mol Microbiol*, **5**(9):2099:2105.

Ochi K. (1986). "*A decrease in GTP content is associated with aerial mycelium formation in Streptomyces MA406-A-1*." *J Gen Microbiol*, **132**(2):299-305.

Oki T, Yoshimoto A, Sato S, Takamatsu A. (1975). "*Purine nucleotide pyrophosphotransferase from Streptomyces morookaensis, capable of synthesizing pppApp and pppGpp*." *Biochim Biophys Acta*, **410**(2):262-272.

Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, Shinose M, Takahashi Y, Horikawa H, Nakazawa H, Osonoe T *et al*. (2001). "*Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites*." *Proc Natl Acad Sci U S A*, **98**(21):12215-12220.

Patte JC, Akrim M, Mejean V. (1998). "*The leader sequence of the Escherichia coli lysC gene is involved in the regulation of LysC synthesis*." *FEMS Microbiol Lett*, **169**(1):165-170.

Reddy PS, Raghavan A, Chatterji D. (1995). "*Evidence for a ppGpp-binding site on Escherichia coli RNA polymerase: proximity relationship with the rifampicin-binding domain*." *Mol Microbiol*, **15**(2):255-265.

Riesenberg D, Bergter F, Kari C. (1984). "*Effect of serine hydroxamate and methyl alpha-D-glucopyranoside treatment on nucleoside polyphosphate pools, RNA and protein accumulation in Streptomyces hygroscopicus*." *J Gen Microbiol*, **130**(10):2549-2558.

Sands MK, Roberts RB. (1952). "*The effects of a tryptophan-histidine deficiency in a mutant of Escherichia coli*." *J Bacteriol*, **63**(4):505-511.

Schreiber G, Metzger S, Aizenman E, Roza S, Cashel M, Glaser G. (1991). "*Overexpression of the relA gene in Escherichia coli*." *J Biol Chem*, **266**(6):3760-3767.

Strauch E, Takano E, Baylis HA, Bibb MJ. (1991). "*The stringent response in Streptomyces coelicolor A3(2)*." *Mol Microbiol*, **5**(2):289-298.

Takano E, Gramajo HC, Strauch E, Andres N, White J, Bibb MJ. (1992). "*Transcriptional regulation of the redD transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in Streptomyces coelicolor A3(2)*." *Mol Microbiol*, **6**(19):2797-2804.

Takano E, Nihira T, Hara Y, Jones JJ, Gershater CJ, Yamada Y, Bibb M. (2000). "*Purification and structural determination of SCB1, a gamma-butyrolactone that elicits antibiotic production in Streptomyces coelicolor A3(2)*." *J Biol Chem*, **275**(15):11010-11016.

Takano E, Chakraburtty R, Nihira T, Yamada Y, Bibb MJ. (2001). "*A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in Streptomyces coelicolor A3(2)*." *Mol Microbiol*, **41**(5):1015-1028.

Takahashi K, Kasai K, Ochi K. (2004). *"Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppGpp) in plants." Proc Natl Acad Sci U S A*, **101**(12):4320-4324.

Tedin K, Bremer H. (1992). "*Toxic effects of high levels of ppGpp in Escherichia coli are relieved by rpoB mutations*." *J Biol Chem*, **267**(4):2337-2344.

Toulokhonov II, Shulgina I, Hernandez VJ. (2001). "*Binding of the transcription effector ppGpp to Escherichia coli RNA polymerase is allosteric, modular, and occurs near the N terminus of the beta'-subunit*." *J Biol Chem*, **276**(2):1220-1225.

Vinciotti V, Khanin R, D'Alimonte D, Liu X, Cattini N, Hotchkiss G, Bucca G, O DJ, Rasaiyaah J, Smith CP *et al*. (2005). "*An experimental evaluation of a loop versus a reference design for two-channel microarrays*." *Bioinformatics*, **21**(4):492-501.

Xiao H, Kalman M, Ikehara K, Zemel S, Glaser G, Cashel M. (1991). "*Residual guanosine 3',5'-bispyrophosphate synthetic activity of relA null mutants can be eliminated by spoT null mutations*." *J Biol Chem*, **266**(9):5980-5990.

Zhou YN, Jin DJ. (1998). "*The rpoB mutants destabilizing initiation complexes at stringently controlled promoters behave like "stringent" RNA polymerases in Escherichia coli*." *Proc Natl Acad Sci U S A*, **95**(6):2908-2913.

CHAPTER 7

Metabolic engineering of *Streptomyces*

Irina Borodina, Anna Eliasson Lantz, and Jens Nielsen

Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, building 223, DK-2800 Kgs. Lyngby, Denmark

The content of this chapter will be published in: Metabolic Pathway Engineering Handbook, 2007, Taylor and Francis Group, LLC

7.1 *Streptomyces* as superhosts

Some of the common soil microorganisms are actinomycetes, Gram-positive bacteria with high GC content. Because of their mycelial habit they were initially believed to be fungi, which reflected in their name (*mucus* (*lat.*) means fungus). In 1939, one year before rediscovery of penicillin by Florey and Chain, soil microbiologist Waksman has set his lab on a quest for new antimicrobial drugs. From the previous studies he knew that actinomycetes can inhibit the growth of other soil bacteria through secretion of bioactive compounds, which he named "antibiotics" (*anti* (*lat.*) against, *bio* (*lat.*) life). Systematic search for antibiotics produced by actinomycetes resulted in the discovery of actinomycin (1940), clavacin and streptothricin (1942), all of them sadly turned out to be toxic in animal tests. In 1943 Waksman's student Schatz isolated streptomycin-producing strain of *Streptomyces griseus*.[1] Streptomycin was not particularly toxic to animals and humans, but remarkably was the first compound active against tuberculosis bacteria. Many pharmaceutical companies and research laboratories started to collect soil samples from all over the world in search of antibiotics-producing organisms. Most of the discoveries were made in the first ten years of the "hunt", the larger part involved *Streptomyces* species. *Streptomyces* is a genus in the genera of actinomycetes, many of these bacteria produce volatile compounds that give the earth its characteristic odor. *Streptomyces* proved to be an excellent source of secondary metabolites, including antibiotics, anticancerous agents, antihelmintic drugs and other useful compounds (Table 7-1). At present more than half of antibiotics in clinical use are produced in *Streptomyces* species.

Table 7-1. Examples of industrial processes that use *Streptomyces*

Product Type	Product	*Streptomyces* sp.	Companies*
Antibacterial	Rifampicin	*S. mediterranei*	Sanofi-Aventis (France)
Antifungal	Nystatin	*S. noursei*	Bristol-Myers Squibb (USA), Bayer (Germany)
Antihelmintic	Ivermectin	*S. averimitilis*	Merck
Herbicidal	Bialaphos	*S. hygroscopicus*	Meiji Seika Kaisha, Ltd (Japan)
Insecticidal	Avermectins	*S. averimitilis*	Novartis (Switzerland)
Antitumour	Bleomycin	*S. verticillus*	Bristol-Myers Squibb (USA)
Immunosupressants	Tacrolimus	*S. tsukubaensis*	Fujisawa (Japan)
Enzymes	Glucose isomerase	*S. murinus*	Novozymes (Denmark)
Heterologous proteins	Human Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF)	*S. lividans* expressing rhGM-CSF gene on a plasmid	Cangene (Canada)

* Only few examples of the producing companies are given. For most of the drugs the original patents have expired and multiple generic variants are available.

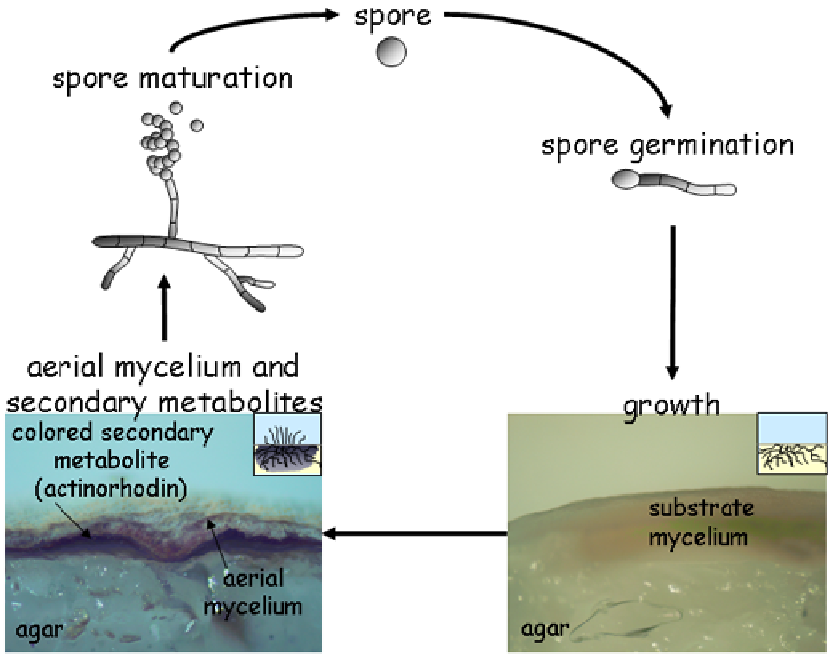
Besides secondary metabolites biosynthesis, *Streptomyces* have also been exploited for enzymes production. Namely *Streptomyces* spp. are capable of decomposing cellulose, lignin, chitin and other complex polymers thanks to excretion of efficient

hydrolyzing enzymes. These enzymes can be applied for degradation of cellulose in bioethanol production, for making high fructose corn syrup, and other processes.[2-4] Some streptomycetes can degrade toxic compounds and are used in bioremediation.[5-8] Due to their ability of protein secretion *Streptomyces* have recently also been used for manufacturing of pharmaceutical recombinant proteins.[9, 10] So *Streptomyces* besides being indispensable in nature as decomposers, also are of high industrial importance.

Very few *Streptomyces* strains are pathogenic, one of them being *S. scabies*, causing scab on tubers of potatoes, beets, etc. Most of the *Streptomyces* strains are nonpathogenic and nontoxic and some of the products made in *Streptomyces* fermentations have a GRAS status (”generally regarded as safe”), e.g., glucose isomerase made by *S. rubiginosus* (Food Standards Agency, UK).

Typical features of *Streptomyces* are mycelial growth and morphological differentiation, which makes them one of the most sophisticated groups of bacteria. In favorable conditions *Streptomyces* spores germinate and produce vegetative mycelium (hyphae); when the nutrients become sparse or the cells get stressed in another way, morphological differentiation begins. Parts of mycelial colonies undergo lysis and hereby nutrients are supplied for the growth of aerial mycelium, which later segments into spores (Figure 7-1). The spores are unicellular with a single copy of the genome, which is in contrast to the hyphae that consist of cells with several nucleoids. *Streptomyces* spores are resistant to the absence of nutrients and dry environments. It is anticipated that many *Streptomyces* strains survive in the soil in the form of spores.[11] During transition from vegetative growth to aerial mycelium formation, *Streptomyces* often secrete bioactive compounds, e.g. antibiotics. Keith Chater has proposed a theory that *Streptomyces* are quite vulnerable at the stage of differentiation and that antibiotics production has been developed as a defense mechanism preventing other soil bacteria to use the nutrients that are released during differentiation.[11] Indeed, about half of the *Streptomyces* strains isolated in the laboratory proved to produce an antimicrobial compound of one or another kind, and many strains produced several antibiotics. Sequencing of *Streptomyces* genomes revealed many secondary metabolite gene clusters per genome, but quite a few of these clusters seem to be silenced under laboratory conditions.[12] *Streptomyces* are the cell factories of choice for production of many secondary metabolites, particularly antibiotics. They are easy to grow in submerged cultures, they can utilize cheap complex industrial media and they secrete metabolites into the broth. It is most common for industry to optimize natural *Streptomyces* isolates employing random mutagenesis and screening and to use the resulting overproducing mutants in production. The strains optimized in this way are not considered to be genetically modified organisms (GMOs), which easies the regulatory requirements around production facilities and the process.

Figure 7-1. Streptomyces life cycle. The pictures show cross-sections of *S. coelicolor* colonies on agar medium. In the upper right corner of each picture the development stage is drawn schematically.



As for production of enzymes and heterologous proteins, *Streptomyces* is one of the hosts to consider. In general, bacterial processes are usually preferred for protein production compared with eukaryotic expression systems, because they are cheaper and easier to run in reactors and give higher yields. In comparison to the typical bacterial host *Escherichia coli*, which accumulates proteins intracellularly in insoluble form, Gram-positive bacteria as *Bacillus subtilis* and *Streptomyces* can efficiently secrete proteins into the medium. They can also express genes with a wide range of GC content without codon adjustment, while *E. coli* has difficulties with high-GC genes expression. The folding of proteins differs in *B. subtilis* and streptomycetes and some proteins are better expressed by the latter. For instance, a large xylogluconase protein from *Jonesia* sp. was successfully secreted in functional soluble form by an engineered *S. lividans*, while attempts to express it in *E. coli* and *B. subtilis* failed.[13] *Streptomyces* are also less susceptible to phage infections than *E. coli* and *B. subtilis*.

On the other hand, there are also disadvantages of using *Streptomyces* instead of classical bacterial hosts as *E. coli* and *B. subtilis*, e.g.:

- mycelial growth limits mass transfer and complicates mixing and oxygen supply;
- tools for genetic manipulation are not as well developed;
- *Streptomyces* genomes are usually very large with many regulatory proteins, which makes it difficult to predict an outcome of a genetic modification.

All in all, *Streptomyces* are very suitable for secondary metabolites production and as more knowledge is accumulated they are also becoming more attractive hosts for production of other compounds.

In this chapter we will briefly discuss *Streptomyces* genomic organization and tools available for genetic modifications. We will thereafter describe the application of some analysis tools that allow more advanced knowledge-based strain design, like microarrays, proteomics, fluxomics, etc. The chapter will end with a review of various strategies that have been used to enhance production of secondary metabolites and proteins in *Streptomyces* species.

7.2. The *Streptomyces* genome and its modification

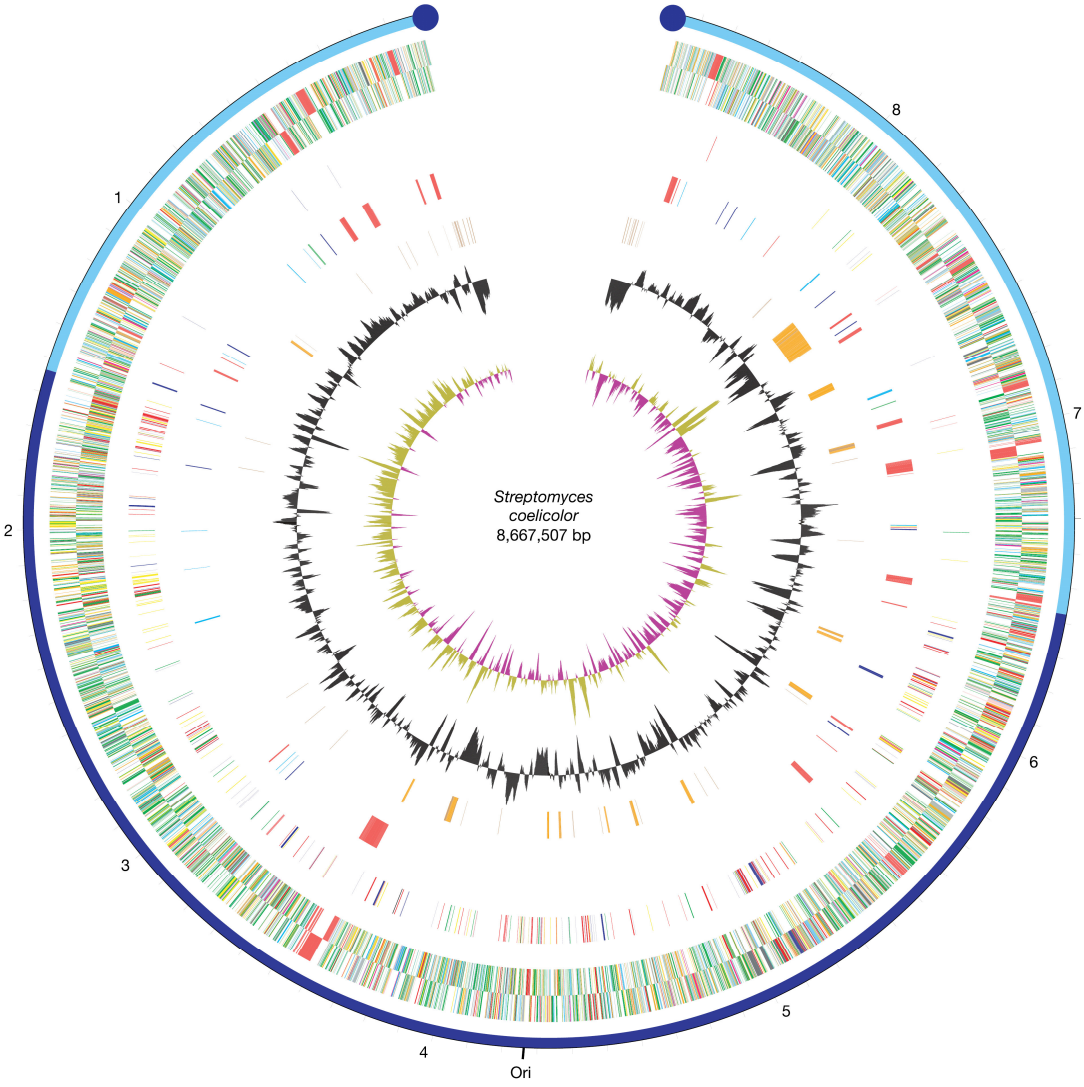
7.2.1 Streptomyces genomes

It is an implication for metabolic engineering that genetic tools for changing gene expression are available. Knowing the genome sequence of the organism gives a tremendous advantage and improves both the speed and the quality of genetic modifications. Until 2002 only sequences of single *Streptomyces* genes or clusters were available. In 1999 Hopwood and co-workers from the John Innes Centre and the Sanger Institute started an initiative on sequencing the whole genome of *S. coelicolor* A3(2), the best genetically studied *Streptomyces* strain, often used as a model antibiotics producer. In 2002 the genome sequence was published in Nature and this marked a new era for *Streptomyces* research.[14] At the time of publication it was the largest bacterial genome sequenced with 8.7 Mb base pairs (Figure 7-2). The coding density was high as also found for other bacteria, and about 8,000 open reading frames were predicted. Thus the bacterium contained more genes than the eukaryotic yeast *Saccharomyces cerevisiae*. The chromosomes of most of *Streptomyces* spp. are linear in contrast to more common circular chromosomes in prokaryotes.[15, 16] To the ends of the chromosome terminal proteins are covalently bound, these proteins probably act as primers for the replication of the last fragment of the lagging strand when the transcriptase reaches the end of the chromosome. The ends contain long terminal inverted repeats. The origin of replication is often positioned around the center of the chromosome. The genes essential for growth are mostly allocated on the core of the chromosome, while the arms contain genes for secondary metabolism and alternative nutrients utilization. The arms of a chromosome are often subjected to mutations, which can cause deletions of up to 2 kb of sequence, duplications of parts of the sequence or circularization of the chromosome[17], consequently the industrial producers are often genetically unstable, that is the product titers can decrease in the course of propagation or cultivation due to spontaneous genetic rearrangements. The genes for secondary metabolism are grouped into clusters, which usually include biosynthetic, resistance and regulatory genes. In the *S. coelicolor* genome 20 clusters were predicted, of which only 3 were known prior to the genome sequencing. It was predicted that around 1,000 proteins play a regulatory role as transcription factors, response regulators, σ–factors, etc. This renders *Streptomyces* one of the most sophisticatedly regulated bacteria and thus makes metabolic engineering of this organism more challenging.

A year after publication of the *S. coelicolor* sequence, the genome sequence of *S. avermitilis*, an industrial producer of the antiparasitic compounds avermectins, was released.[18] The sequence of the plant parasite *S. scabies* has become available in 2006, although the annotation of the genome was not yet completed at the time of writing. Other *Streptomyces* genomes are in the pipeline for genomic sequencing: *S. ambofaciens*, *S. peucetius* (Genomes OnLine Database, www.genomesonline.org).

S. diversa genomic sequencing was performed by Diversa/Celera but the sequence remains proprietary so far.

Figure 7-2. *Streptomyces coelicolor* chromosome. The outer scale is numbered anticlockwise in megabases and indicates the core (dark blue) and arm (light blue) regions of the chromosome. Circles 1 and 2 (from the outside in), all genes (reverse and forward strand, respectively) color-coded by function (black, energy metabolism; red, information transfer and secondary metabolism; dark green, surface associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central or intermediary metabolism; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pale green, unknown; grey, miscellaneous); circle 3, selected 'essential' genes (for cell division, DNA replication, transcription, translation and amino-acid biosynthesis, color coding as for circles 1 and 2); circle 4, selected 'contingency' genes (red, secondary metabolism; pale blue, exoenzymes; dark blue, conservon; green, gas vesicle proteins); circle 5, mobile elements (brown, transposases; orange, putative laterally acquired genes); circle 6, G + C content; circle 7, GC bias ((G - C/G + C), khaki indicates values >1, purple <1). The origin of replication (Ori) and terminal protein (blue circles) are also indicated.[14]



7.2.2 Molecular biology of Streptomyces

Molecular biology work with *Streptomyces* has its own rules and tricks, which are well described in an excellent handbook “Practical *Streptomyces* Genetics” by Kieser et al. from the John Innes Center.[11] Here we will just briefly go through the main difficulties that one might stumble into in the beginning of working with *Streptomyces*. *Streptomyces* have a very high GC content, e.g. 72% in the *S. coelicolor* genome. This complicates PCR reactions, sequencing and might render plasmids unstable in *E. coli*. Furthermore many *Streptomyces* have a restriction barrier, which makes it impossible to transform them with methylated DNA.[19] To avoid methylation the plasmids can be passed through non-methylating *E. coli* strains (as ET12567)[19] or through *S. lividans* before introduction into the host. It is common to use conjugation from *E. coli* to deliver the DNA into *Streptomyces* and the method works well with nearly all species.[20, 21] Some strains can be transformed by PEG-assisted protoplast transformation or transfected by phages. One should be careful when choosing antibiotics selection markers because many strains will be naturally resistant to certain antibiotics.

7.2.2.1 Overexpression of genes

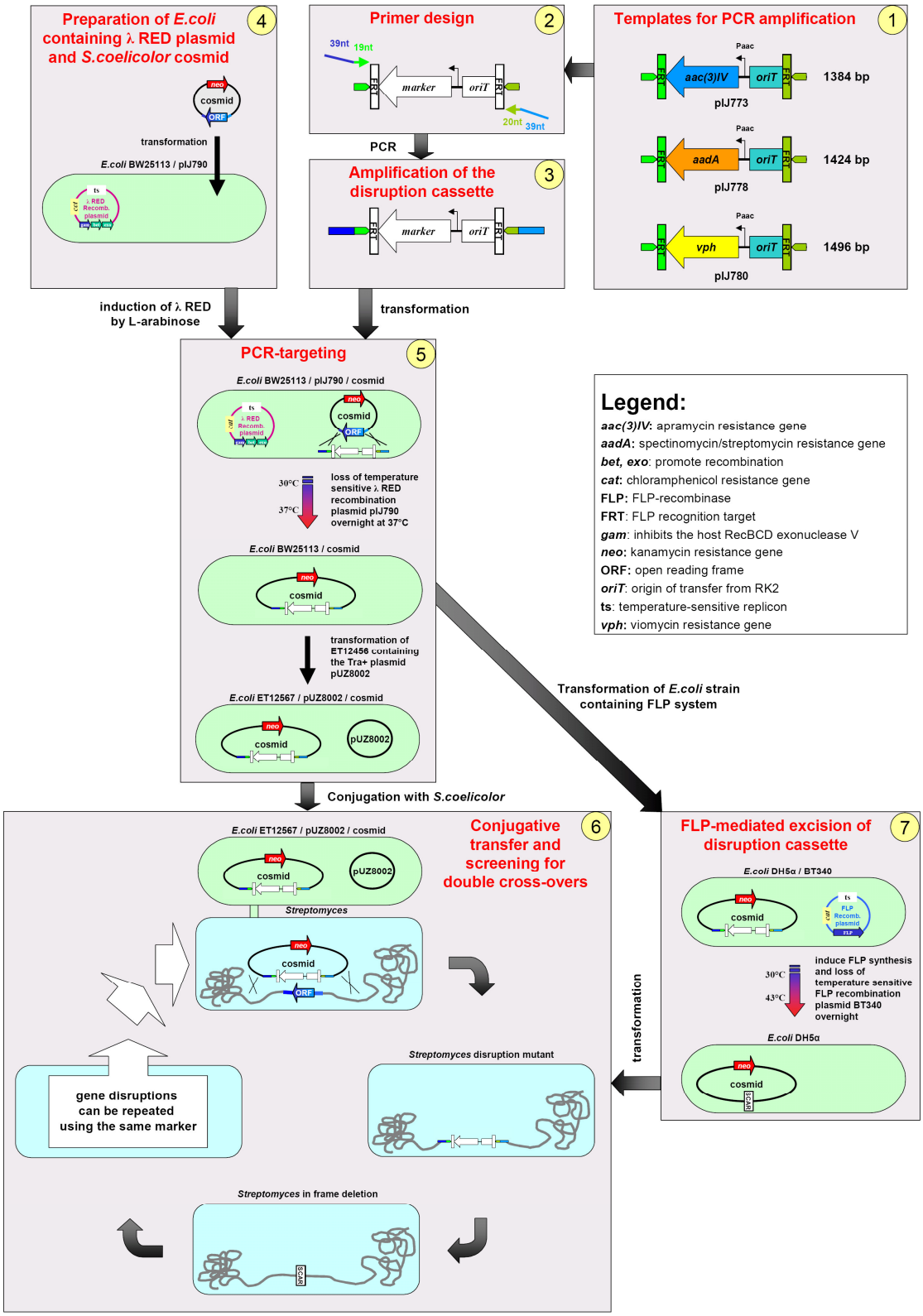
Many vectors for over-expression of genes in *Streptomyces* have been published and patented recently. These include both self-replicating and integrating vectors. There have been some reports that multicopy vectors reduce antibiotics production, so it is advisable to use integrating vectors whenever possible. High-producing strains seem to be particularly sensitive to the metabolic load imposed by multicopy plasmids and vectors are therefore lost in the absence of selection pressure.[11] Integrative vectors often contain the attachment site attP and the gene *int*, coding for integrase from phage phiC31.[22, 23] Many *Streptomyces* carry a attB sequence, identical to attP, in their chromosome. The integrase protein promotes recombination between att sites and thus causes vector integration. Even when there is no attB sequence present in the chromosome, integration can still occur through recombination with sequences, which are homologous to attB.[24] A range of strong constitutive and inducible promoters are available, e.g. thiostrepton-inducible tipA[25], modified constitutive ermP* [personal communication with Mervyn Bibb, John Innes Centre, UK], and snpR-activated snpA[26, 27]. To achieve a range of expression levels one can also use synthetic promoter libraries.[28, 29]

7.2.2.2 Replacement of genes

The classical gene inactivation method works by interrupting the desirable gene by a plasmid through homologous recombination. Depending on the number and position of homologous sites on the plasmid single or double crossover occurs and the gene is either interrupted by insertion of a foreign fragment or is partly or totally substituted by a selection marker. The selection marker can be afterwards removed leaving a small nucleotide sequence, called “scar”, in the place of the deleted gene. The process of plasmid construction and mutant selection is normally quite time-consuming and can easily take several man-months. The real breakthrough for yeast genetics was the discovery of a PCR-based gene replacement method[30], which reduced the time needed to make a mutation down to a few days and furthermore it could be almost fully automated. Later they found out that a similar strategy could be applied in *E. coli* by expressing λ-recombinase from a low-copy number easily curable plasmid.[31] The success with *E. coli* has paved the way for a faster gene replacement method for *Streptomyces.*[32] The method, called REDIRECT, is based on replacement of a gene on a cosmid by a PCR fragment in an *E. coli* strain, engineered for efficient λ-recombination and for low degradation of linear DNA strands. The mutated cosmid is then extracted from *E. coli* and introduced into *Streptomyces*, where it by homologous recombination replaces the gene of interest. In comparison to the plasmid-based methods, the mutant cosmid construction is quite fast and the success of recombination in *Streptomyces* is increased due to very large homologous regions on the cosmid. It is possible to loop out the resistance cassette later, leaving a “scar”, which will not disturb the transcription of the surrounding genes. This allows studying an effect of single gene deletion even when the gene is a part of an operon.[32] The removal can either be done using FLP recombination or Cre-loxP system. In FLP recombination method the mutated cosmid is transferred into an *E. coli* strain carrying FLP recombinase on a plasmid. After the recombination has occurred the cosmid with the “scar” can be isolated from the *E. coli* and transferred into the mutated *Streptomyces* strain. The replacement of resistant marker by “scar” in *Streptomyces* is selected by loss of antibiotics resistance.[32] Another procedure suggests to use loxP sequences instead of FLP

recombination targets and to use an engineered Cre phage to infect the mutated *Streptomyces* strain and to perform the loxP recombination directly inside the *Streptomyces* cell. The phage infection can be cured in a few sub-cultivations.[33]
The prerequisite for using this knock-out method is the availability of a cosmid with the gene of interest, which is not always readily available for industrial strains.

Figure 7-3. Flowchart of gene disruption in *Streptomyces* using PCR targeting.[32] 1 – The knock-out cassette, containing antibiotics resistance marker, origin of transfer and elements for cassette loop-out, is cut from the plasmid and purified. 2, 3 – The cassette is amplified by PCR using primers which carry upstream and downstream sequences of the gene which should be replaced. 4 – *Streptomyces* cosmid is transformed into an *E. coli* strain BW25113/pIJ790, which was designed for improved homologous recombination. 5 – The *E. coli* carrying the cosmid with the gene of interest is transformed with the knock-out cassette. On induction of λ-RED genes by arabinose, the cassette integrates into the cosmid by homologous recombination. The λ-RED plasmid with a temperature sensitive replicon is cured by inoculation at 37ºC. The mutated cosmid is isolated and transformed into a non-methylating *E. coli* strain ET12567, which carries transfer functions on plasmid pUZ8002. 6 – The mutated cosmid is transferred into *Streptomyces* by conjugation from *E. coli* and the mutants screened, the double cross-over transformants should be resistant to the chosen resistance marker and sensitive to kanamycin. 7 – The knock-out cassette can be removed from the mutated cosmid in *E. coli* with expressed FLP-recombinase gene and the newly made cosmid with a “scar” can be introduced into the mutated *Streptomyces* strain. The double cross-over event is selected by loss of antibiotics resistance.


Templates for PCR amplification
1
Paac
aac(3)IV
oriT
FRT
1384 bp
pIJ773
aadA
1424 bp
pIJ778
vph
1496 bp
pIJ780
Primer design
2
39nt
19nt
marker
20nt
39nt
PCR
Amplification of the disruption cassette
3
Preparation of E.coli containing λ RED plasmid and S.coelicolor cosmid
4
neo
cosmid
ORF
transformation
E.coli BW25113 / pIJ790
induction of λ RED by L-arabinose
transformation
PCR-targeting
5
E.coli BW25113 / pIJ790 / cosmid
30°C
37°C
loss of temperature sensitive λ RED recombination plasmid pIJ790 overnight at 37°C
E.coli BW25113 / cosmid
transformation of ET12456 containing the Tra+ plasmid pUZ8002
E.coli ET12567 / pUZ8002 / cosmid
pUZ8002
Legend:
aac(3)IV: apramycin resistance gene
aadA: spectinomycin/streptomycin resistance gene
bet, exo: promote recombination
cat: chloramphenicol resistance gene
FLP: FLP-recombinase
FRT: FLP recognition target
gam: inhibits the host RecBCD exonuclease V
neo: kanamycin resistance gene
ORF: open reading frame
oriT: origin of transfer from RK2
ts: temperature-sensitive replicon
vph: viomycin resistance gene
Transformation of E.coli strain containing FLP system
Conjugation with S.coelicolor
E.coli ET12567 / pUZ8002 / cosmid
Conjugative transfer and screening for double cross-overs
6
Streptomyces
gene disruptions can be repeated using the same marker
Streptomyces disruption mutant
Streptomyces in frame deletion
FLP-mediated excision of disruption cassette
7
E.coli DH5α / BT340
FLP Recomb. plasmid
induce FLP synthesis and loss of temperature sensitive FLP recombination plasmid BT340 overnight
30°C
43°C
E.coli DH5α
transformation

7.2.2.3 Transposon mutagenesis

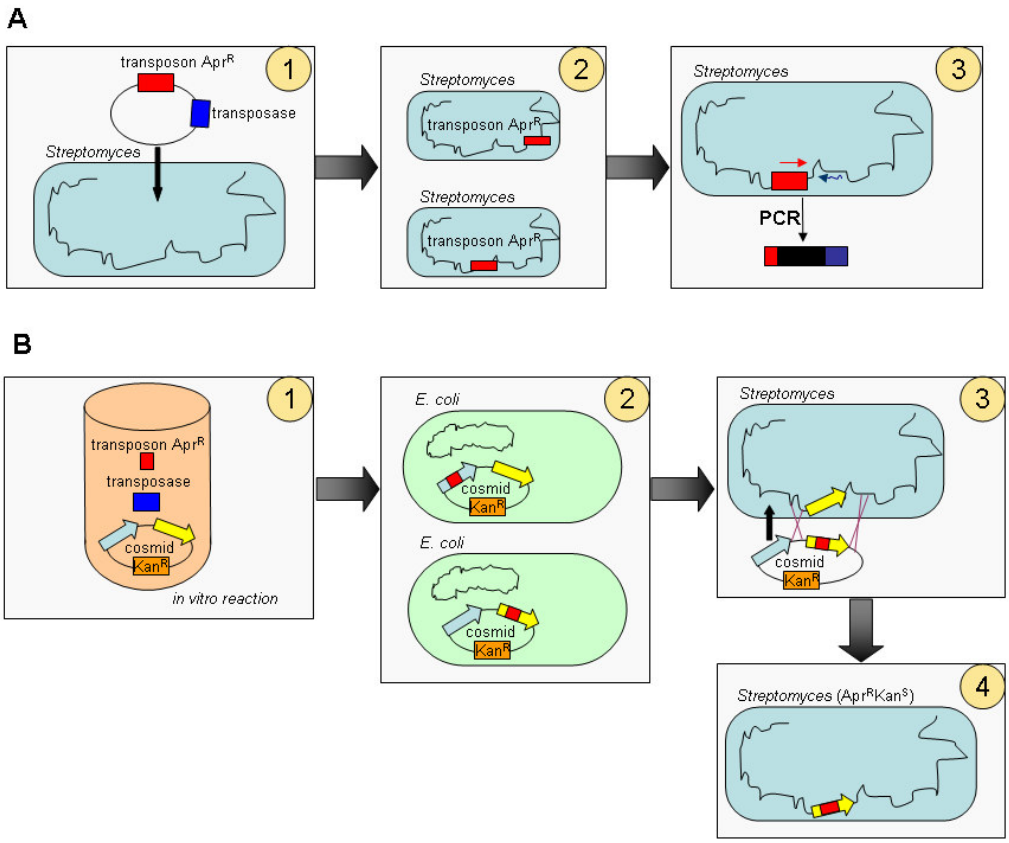
Genome-scale mutant libraries are interesting for systems biology approaches, for example to relate phenotypic features to the genotype. To generate large libraries, transposon-based mutagenesis techniques have been successfully applied.[34-36] Transposons are mobile genetic elements, which can change their position in the genome. Depending on their structure they are divided into class I and class II transposons. Class I transposons are composed of some gene (e.g., drug resistance gene) flanked by insertion sequences (IS). ISs are typically 0.7-1.8 kb long sequences with repeats at the termini; they code for a single protein, which is involved in transposition. Class I transposons can transpose as a unit, but the flanking ISs can also transpose separately. Class II transposons carry a transposase and some other gene (-s) flanked by indirect repeats of 30-40 bp, they can only transpose as the whole unit. Genetically altered transposons lacking transposase have been designed. Transposase can be supplied *in trans* for the transposition event and then removed; without transposase the transposon insertion is stable in the genome.

There are some difficulties associated with transposon mutagenesis. More than one insertion per genome can occur and hence some mutants can carry insertions in several places of DNA. The problem can be circumvented by using Tn3-like transposon, which insertion results in transposition immunity whereby a second insertion in the DNA molecule is inhibited.[37] Another problem is that transposons are usually biased towards integrating in particular sequences and so the obtained mutations will not be evenly distributed over the chromosome.[11, 38] Finally, the transposons are usually polar, that is upon insertion they will disturb expression of the downstream genes. This is because the flanking regions of the transposons normally serve as termination signals for RNA transcription and transposons have nonsense codons in all reading frames causing translation termination.

In vivo transposon mutagenesis is performed by introducing transposase on a suicide plasmid and a transposon into the host and selecting for antibiotics resistant phenotype caused by transposon insertion into the chromosome (Figure 7-4A). The result of *in vivo* mutagenesis is a mix of mutants, which can further be isolated and characterized to find out which gene that has been disrupted. Because identification of insertion position is at present a time-consuming procedure, only strains with interesting phenotypes are generally investigated. A mutant library of *S. lividans* has been generated by *in vivo* transposon mutagenesis.[39] Screening identified a bald mutant, which was investigated further. By inverse PCR and sequencing it was determined that the mutant had an insertion in the osmoadaptation regulator *osaB* gene. The mutant turned out to overproduce antibiotics on medium with high osmolytes concentration.[40] *In vitro* mutagenesis can be used as a targeted approach. Here the mutagenesis can be performed by mixing the cosmids containing the sequence of the host organism with the transposon and purified transposase (Figure 7-4B). The mutagenized cosmids are transformed into *E. coli*, isolated and the insertion position is determined. To obtain a desired insertion, a suitable cosmid is transformed into the host, where its integration by double cross-over can be detected using selection markers. A collection of *S. coelicolor* mutagenized cosmids is being constructed with the funding from The Biotechnology and Biological Sciences Research Council (BBSRC, UK) and is available for academic research from Swansea University, UK. The positions of the insertions are shown on the SCODB website (http://streptomyces.org.uk/sco/index.html). At the moment of writing the

library contained mutagenized cosmids for disruption of about 60% of the open reading frames. In the cosmids that have been processed single insertions were obtained for about 90% of the open reading frames; the rest of the ORFs will have to be inactivated by a different method (e.g., REDIRECT).

Figure 7-4. Simplified scheme of *in vivo* (A) and *in vitro* (B) transposon mutagenesis. A. 1 –*Streptomyces* is transformed with a suicide plasmid containing transposon and transposase gene. 2 – *Streptomyces* insertion mutants are selected by apramycin resistance. 3 – Mutants with interesting phenotypes are characterized by PCR to determine the transposon insertion position. B. 1 – Cosmid containing a part of *Streptomyces* genome is mixed with transposon and purified transposase *in vitro*, the resulting mutated cosmids are electroporated into *E. coli*. 2 – Antibiotics resistant clones are selected; the position of insertions is characterized by PCR for a large number of clones. 3 – For disruption of a particular gene (yellow) a suitable mutated cosmid is transformed into *Streptomyces* by conjugation from *E. coli*. 4 – *Streptomyces* colonies that are apramycin resistant and kanamycin sensitive are selected; they result from double cross-over event and carry insertion in the gene of interest.



The disadvantage of gene inactivation by insertion instead of deletion is that: firstly, the gene can remain partly or fully functional with insertion, particularly if the insertion is further from the 3´-end of the gene[41, 42] and secondly, the expression of the downstream genes will most likely be disturbed by transposon, making the method unsuitable for studying knock-outs of genes, which are parts of an operon.

7.3 Analysis of *Streptomyces* strains

During the past years we have witnessed an extensive development of analytical techniques in biological sciences. They have allowed quantitative measurements of various cellular components from RNA and proteins to metabolites. The advances in analytical methods along with progress in computer driven analysis of data gave birth to systems biology, where the cells are viewed as a system of various interacting components. One outcome of systems biology is integrative models that allow us to make better estimations of how we can change the cellular metabolism in a desired direction, so we can make metabolic engineering with a higher success rate.

7.3.1 Transcriptome

As soon as the genomic sequence of *S. coelicolor* started to be released, the first cDNA arrays appeared. All of the transcriptome work published until today are done with PCR – or oligonucleotide arrays available for academic research from Stanford School of Biomedical and Molecular Sciences (USA) and from Surrey University (UK). It is also possible to buy on-situ synthesized arrays (NimbleGen Systems Inc., USA), though the high price is a barrier for their extensive use.

Gene expression analyses have cast light on many transcriptionally regulated processes in *Streptomyces*. In the first published study the authors monitored gene expression during the course of a batch fermentation of *S. coelicolor*. They observed that expression of secondary metabolite gene clusters as well as expression of many biosynthetic, ribosomal and regulatory genes was changing during the different developmental phases. This paper also presented an algorithm for finding boundaries of clusters of co-expressed genes.[43] In another study the same group tested the hypothesis that the genes in the central part of the chromosome encode "core" functions connected to growth. They found that during the non-limited growth indeed the expression level was higher in the core of the chromosome than on its right and left arms. During the stationary phase and at different stress conditions the expression of the core genes decreased and various other genes throughout the chromosome composed the bigger part of the transcripts. Those were genes related to stress, morphological differentiation, and other less characterized genes.[44]

There has also been a few studies on regulatory mutants of *S. coelicolor*, e.g. with deletions of specific activators of antibiotics gene clusters (*actII-4* and *redD*)[43], deletion of the heat shock protein *hspR*[45] and deletion of the two-component element *absA1* with an effect on antibiotics genes expression[46].

A range of bioinformatic tools have been used for prediction of operons in *S. coelicolor.*[47] When expression data from some of the previous studies was overlaid on the predicted operons, it was observed that expression usually decreases from the first to the last gene in a given operon. This polarity of gene expression is not observed in *E. coli* and is thought to be attributed to the *Streptomyces* high GC content, which complicates the transcription process. A few genes positioned in the middle of the operon were expressed higher than the first gene, but in front of those genes internal promoters were commonly found.[47] The global transcriptional studies have revealed regional fluctuations of gene expression across the chromosome (Colin Smith, Surrey University, UK, presentation at Genetics of Industrial Microorganisms 2006). One of the explanations could be the variation in the binding of histone-like proteins to DNA. The *S. coelicolor* histone-like proteins are small proteins around 12-16 KDa, which bind to the chromosome in form of dimers or heterodimers, making the nucleoid. It's been proven that genes transcription was

positively correlated with their exposure as determined by sensitivity to DNase I treatment.[48] Hopefully further research in gene expression will provide system-level knowledge on transcriptional regulation in *Streptomyces*.
Microarrays can also be used to analyze overproducing mutants generated by random mutagenesis in order to understand why overproduction occurs and use the knowledge for directed design of strains, in the process of inverse metabolic engineering. The industrial overproducers of erythromycin and tylosin, *Saccharopolyspora erythraea* and *S. fradiae* respectively, were compared to the corresponding wild type strains using spotted arrays. The arrays were based on the *S. coelicolor* genomic sequence but additionally contained spots for the erythromycin and tylosin gene clusters. The industrial *S. erythraea* was found to have longer and more extensive expression of antibiotics biosynthetic genes than the wild type, while the industrial *S. fradiae* strain had changes in the expression of some metabolic genes involved in biosynthesis of the tylosin precursor.[49]

7.3.2 Proteome

The levels of mRNA, which can be measured by microarrays, do not necessarily correlate with protein concentrations. Besides, protein activity, function and half-life time are greatly affected by posttranslational modifications. Protein concentrations and modifications can be estimated in proteomics analysis and the information may be helpful in metabolic engineering.
Hesketh et al. reported results of 2D gel electrophoresis of *S. coelicolor* whole cell extracts followed by MALDI-TOF mass spectrometry.[50] About 10% of all the predicted *S. coelicolor* proteins were detected. Positively, the sequences of all the proteins agreed with previously predicted sequences,[14] signifying the high quality of the genome annotation. On average 1.2 spots were detected per protein sequence, which shows that there is a high degree of posttranslational modifications. 2D gels provide more qualitative than quantitative information, which limits data interpretation. A more advanced approach to proteome analysis is multidimensional LC-LC-MS, which has been applied for analysis of the industrial *S. diversa* strain (Steve Briggs, Diversa, USA, presentation at the conference Metabolic Engineering V).

7.3.3 Metabolome

There are no published data on metabolome analysis in *Streptomyces*. For the analysis of intracellular metabolites a popular quenching method with cold methanol is not advisable because of the substantial leakage of intracellular metabolites.[51, 52] Rapid filtration followed by quenching with liquid nitrogen can be an alternative[53], but this is not very suitable for metabolites with high turnover rates.

7.3.4 Fluxome

In metabolic engineering comparison of flux maps for reference and mutant strains represents an invaluable source of information about operation of the metabolic network in the two strains. Metabolic fluxes for *Streptomyces* have often been calculated applying metabolic models of various sizes based on only a few measurements of fluxes in and out of the cell.[54-56] One should be careful in interpretation of these data, as these metabolic models often have a high degree of freedom and hence the solution identified, i.e. the set of fluxes, may not be unique.
More sound flux data is generated by applying isotope labeling technique for metabolic flux analysis (^{13}C MFA), where metabolite balancing is supplemented by

isotope balancing making the models more determined. The enrichment of compound with isotopes can be determined by gas chromatography coupled to mass spectrometry (GC-MS) or ^{13}C nuclear magnetic resonance (NMR). Besides enabling robust quantification of the metabolic fluxes this approach may be used to map the network topology. Thus, GC-MS analysis of cells grown on [1-^{13}C]-glucose led to discovery of a glycolytic pathway unusual for actinomycetes in *Nonomuraea* sp. 39727[57] and in *S. tenebrarius*[58]. In both these organisms the Entner-Doudoroff (ED) pathway was identified, and its presence has implications on the reducing cofactor biosynthesis and energy metabolism, because of a different stoichiometry in this pathway compared with the Embden-Meyerhof-Parnas (EMP) pathway. *In vivo* ^{13}C and ^{15}N NMR analysis of actinomycin D-producing *Streptomyces parvulus*, grown on fructose-glutamate medium, revealed high flux via gluconeogenesis as well as the origin of some of actinomycin D precursor amino acids.[59] In other studies using C13 MFA it was possible to spot changes in metabolism occurring over the time course of cultivation when the cells progress from exponential growth to the stationary production phase. In both actinorhodin-producing *S. lividans* and in nystatin-producing *S. noursei*, a decrease in the pentose phosphate (PP) pathway activity was observed upon onset of antibiotics production.[54, 60] In actinomycetes that have an active ED pathway (*Nonomuraea* sp. 39727 and *S. tenebrarius*), the ED pathway flux was found to decrease and more glucose was diverted to the EMP and pentose phosphate (PP) pathways upon onset of production.[58, 61] C13 flux analysis of a *S. coelicolor* mutant with deletion of one of the phosphofructokinase isoenzymes (PFKA2) showed that this mutant had a higher flux through the PP pathway. Pentose phosphate pathway supplies NADPH, a cofactor used in many antibiotics biosynthesis. The increase of PP flux correlated with overproduction of the pigmented antibiotics actinorhodin (ACT) and prodigiosins (RED) in the mutant.[52]

7.4 Modeling and design of *Streptomyces* strains

Whole cellular models still remain only a far-away perspective, even for such well-studied organisms as *E. coli* and yeast. For the complex *Streptomyces* bacteria availability of whole cell models, which include metabolism, regulation, kinetics, etc. is not likely for a very long time. However, for the purpose of metabolic engineering simpler genome-scale metabolic models have proven to be useful.

Genome-scale metabolic models represent most of the metabolic reactions known to be performed by a given organism either from its genomic sequence or from physiological and biochemical data. As an example of genomic evidence, the presence of lactate dehydrogenase in the *S. coelicolor* genome implies that the cells can reduce pyruvate into lactate at the expense of NADH. As an example of physiological evidence could be the ability of *S. coelicolor* to degrade starch, telling that the organism possesses amyloglucosidases. Furthermore, biochemical data tells for instance that the *S. coelicolor* proteins contain histidine and as the cells don't require amino acids supplementation for growth, all the enzymes for histidine biosynthesis must be present.

A genome-scale metabolic model has been reconstructed for *S. coelicolor*[62] and models for a few other *Streptomyces* are under way (e.g., *S. scabies*). The models can be useful for – omic data interpretation[63] and for predictions, e.g. of strain improvement targets[64, 65].

The reporter metabolite methodology,[66] which identifies around which metabolites most pronounced transcriptional changes have occurred, was applied for analysis of

transcriptional data from phosphofructokinase deletion mutant (Δ*pfkA2*) as compared to the reference strain. One of the most scoring reporter metabolites was NADPH, which is consistent with the increased pentose phosphate pathway flux in the mutant.[52]

Model guided strain design is a new phenomena, and hasn't been used for *Streptomyces* yet, however it has been successfully applied for other organisms, e.g. for improvement of lycopene production in *E. coli*.[67, 68] This systematic metabolic engineering approach could allow design of superhosts for production of different classes of secondary metabolites. Another application could be design of strains with certain metabolic problems (like excess production of NADPH). These strains can be used for improvement of enzymes properties. For example, for improvement of NADPH-dependent reductase, it is inserted into an engineered NADPH-overproducing strain and the strain is grown for many generations until the growth rate is improved. The resulting enzyme-coding gene is isolated and it will often carry mutations that have improved enzymatic activity. (METEVOL® technology from METabolic EXplorer, France).

7.5 Examples of metabolic engineering in *Streptomyces*

7.5.1 Using an optimized host

Development of industrial *Streptomyces* production strains usually starts (and often ends) with chemical or UV mutagenesis followed by selection. Another popular method is protoplast fusion, which often results in breeds performing better than the parental strains. These methods are simple and relatively easy to apply; they do not require priory knowledge about the biosynthetic pathway, coding genes, etc. The disadvantages are that they are random, non-reproducible, can take long time and have to be applied to each strain that produces a new product.

Many new secondary metabolite-producing organisms have been discovered recently. To evaluate the bioactivity of new metabolites it is necessary to produce them in fair amounts, which is often a problem with natural producers. Optimizing each wild type strain is tedious and expensive. For secondary metabolite genes, which are isolated from soil samples or generated by gene shuffling and other procedures, there is no strain to start with at all. A good expression host would greatly speed up the testing process and further enable high level production if the metabolite turns out to possess interesting properties.

The most natural solution would be to have a range of optimized hosts for production of compounds from a particular class (β-lactams, polyketides, non-ribosomal peptides, etc.). The hosts should be clean, that is not being able to produce native secondary metabolites in significant amounts. They should also be capable of producing precursors for many compounds belonging to the same class. One of the options is to develop superhosts from already optimized industrial strains. Industrial strains of *S. fradiae* and *Saccharopolyspora erythraea* have been engineered into plug&play superhosts for polyketide production.[69] To start with, the whole native clusters of tylosin and erythromycin were deleted. *S. erythraea* did not contain a specific attB site for phage integration and hence this site was introduced to allow integration of the plasmid delivering biosynthetic genes for a new product. To test the hosts, the biosynthetic genes were then cloned from wild type low-producing strains of *S. fradiae* and *S. erythraea* and introduced into the engineered hosts using integrative plasmids. The levels of antibiotics production did not differ from that

obtained with the endogeneous antibiotic clusters in the original industrial strains, even when native promoters for the endogenous biosynthetic genes were used. It was found that the promoter sequences did not differ in the mutated industrial and the wild-type strains, indicating that during the optimization of these strains most of the changes did not occur in the sequences of the biosynthetic genes or their promoters, but elsewhere. Better production in industrial strains might instead have been caused by increased substrates uptake rates, better production of Gibbs free energy and precursors and most probably also deregulation of secondary metabolite genes expression. One should, however, keep in mind that sometimes the strains can be overproducing for other reasons, like multiplications of the biosynthetic clusters (see also below, section 5.2), and obviously such strains do not represent a suitable superhost platform for production of other antibiotics. Interestingly, expressing the biosynthetic gene clusters behind a strong constitutive promoter like *ermEP** or by using the *S. coelicolor actUIp/actII2-4* expression system in the above mentioned industrial strain backgrounds resulted in 20-fold decrease in antibiotics titers.[69] This is curious because there has been a successful attempt to improve erythromycin production through substitution of the native promoter in a wild type *S. erythraea* to the *actUIp/actII2-4* system.[70] However, in the industrial strains the regulation might be changed and the expression of the biosynthetic genes from the newly introduced promoters might have been weaker than from the original ones.[69] The barrier in using industrial strains as superhost platforms is that they are usually proprietary and not sequenced. Alternatives could be well-studied strains such as *S. coelicolor* and *S. lividans*. Both of these strains have been used for heterologous expression of secondary metabolites biosynthetic genes from other organisms that are poorly characterized genetically or are problematic to grow in the lab. Expression of a large modular polyketide synthase from *S. erythraea* of about 30.8 kb long in *S. coelicolor*[71] and in *S. lividans*[72] gave good yields of 6-deoxyerythronolide B, the aglycone unit of erythromycin. The myxobacterium *Sorangium cellulosum* is a natural producer of the valuable anti-cancer agent epothilone, but the bacterium is slow-growing and forms multicellular fruiting bodies on agar plates, which complicates genetic engineering.[73] The epothilone-coding genes were therefore cloned into *S. coelicolor* CH999 host, which has the actinorhodin biosynthetic genes deleted and the prodigiosins production blocked.[74] This resulted in a strain that could produce up to 0.1 mg/L epothilones A and B in non-optimized fermentation.[75] The titer was lower than in *Sorangium cellulosum* fermentations (up to 20 mg/L), but similar values could be reached after medium optimization. The main advantage of this strategy is that *S. coelicolor* grows ten times faster than the native host and is well amenable for genetic manipulations, which is useful for generating epothilone analogues.

7.5.2 Increasing expression of genes from the biosynthetic cluster

In the wild type strains secondary metabolite genes are often weakly expressed or not expressed at all under the given conditions and the yields of the compounds are very low. Over-expression of the whole cluster or of some of the rate-controlling enzymes will often result in higher production. Multiplication of the whole cluster was observed in industrial strains that have undergone empirical selection. A commercial *S. kanamyceticus* strain (Meiji Seika Kaisha Ltd., Japan) used for kanamycin production contains up to 36 copies of the kanamycin biosynthetic cluster.[76] The cluster is located within an amplifiable unit of DNA (AUD), which made this high copy number possible. AUD consists of internal sequence flanked by direct repeats, AUDs

usually are located in the arms of the chromosome, which are genetically unstable. AUDs can make tandem repeats, which include several hundred copies of the DNA unit.[39] The amplification is often accompanied by deletions of other regions of DNA. Interestingly, Meiji Seika Kaisha's strain is maintained as a mycelial culture because the high producing kanamycin phenotype gets lost in a single sporulation round. The authors hypothesized that the problem lies in the limited space for chromosome packaging in the spore, which can not accommodate the 5 Mb long amplification.[76] Another way to increase gene expression is to replace a native promoter with a stronger, possibly inducible, promoter. Promoter substitution has been used to increase the expression of the xylose isomerase gene (*xylA*). This gene is normally induced by xylose and inhibited by glucose, and in order to have constitutive expression the *xylA* gene was put under control of a constitutive promoter and the construct was integrated on the chromosome of *Streptomyces violaceoniger*. This resulted in four to five fold improved xylose isomerase activity in the absence of xylose as inducer.[77] Higher expression of whole biosynthetic gene clusters has been achieved by overexpressing positive regulators of the pathway or by deleting negative regulators. When the activator of the actinorhodin cluster (*actII-4*) or the activator of the prodigiosin cluster (*redD*) were introduced into *S. coelicolor* on a multicopy plasmid, the production of ACT and RED increased.[78, 79] In another study two polyketide synthase clusters were found by DNA fingerprinting in *Streptomyces* sp. PGA64: rubromycin cluster and a putative augucycline cluster. The second cluster was most probably silent under the given fermentation conditions because the strain produced only rubromycin in detectable amounts. Disruption of a putative augucycline cluster repressor resulted in a partial activation of the cluster and in secretion of two new augucycline metabolites - UWM6 and rabelomycin.[80] Overexpression of one or a few enzymatic genes from the biosynthetic cluster has also been a successful strategy. The challenging part is to determine which enzyme that has high flux control. In the fermentations of *S. fradiae* two major macrolides were found: tylosin, which is the product of interest and the tylosin precursor – macrocin. In the industrial strains from Eli Lilly they reported that tylosin accounted for 50-55% of all macrolides produced, while macrocin accounted for about 40%.[81] This clearly indicated that the conversion of macrocin into tylosin is rate-limiting in these strains. This process is catalyzed by a single enzyme macrocin O-methyltransferase TYLF, and by inserting a second copy of the *tylF* gene it was possible to increase the fraction of tylosin in the macrolide pool from 50 to 80-85%.[81]

7.5.3 Increasing precursor or cofactor supply

Other approaches to enhance secondary metabolite production are to increase the biosynthesis of the precursor metabolites and to remove or hamper pathways that compete for this precursor with the secondary metabolites reactions. For example, malonyl-CoA and methylmalonyl-CoA serve as precursors for many polyketide antibiotics. To escalate the supply of these precursors the dicarboxylate transporter *matC* and malonyl-CoA synthase *matB* from *Rhizobium trifolii* were expressed in *S. coelicolor*, resulting in a strain that could take up relatively cheap substrates as malonate and methylmalonate and convert them into corresponding CoA thioesters. The strategy was tested in *S. coelicolor*, producing macrolactone 6-deoxyerythronolide B, and resulted in a 3-fold increase in the macrolactone titer.[82] Also overexpression of acetyl-CoA carboxylase, which converts acetyl-CoA into malonyl-CoA had a positive effect on production of polyketide compound

actinorhodin in *S. coelicolor*.[83] Nikkomycin production in *S. ansochromogenes* was improved by overexpressing two genes sanU and sanV, which together with coenzyme B12 form an active glutamate mutase, the latter participates in the biosynthesis of a precursor of nikkomycin.[84] A more straightforward approach is to add precursors directly to the fermentation medium, and if the precursor(s) price is low this can be a feasible industrial solution. Thus, it has been found that addition of glycine, arginine, phenylalanine and tyrosine improve vancomycin production in *Amycolatopsis orientalis.*[85] Tyrosine is a direct precursor in the pathway, whereas phenylalanine degradation leads to formation of another precursor 4-hydroxyphenylpyruvate.

Some precursors also lead to induction of the genes encoding enzymes involved in biosynthetic pathways. For example, lysine is a precursor for the first enzyme of cephalosporin biosynthetic pathway (L-Lysine epsilon-aminotransferase (LAT)), but it also induces expression of this enzyme in *S. clavuligerus*, and addition of lysine to defined medium improved cephamycin production.[86]An example of removing a competing pathway is blocking of clavams and cephamycin C biosynthesis in industrial *S. clavuligerus*, which improved clavulanic acid titers by 10%.[87] Disruption of glyceraldehydes-3-phosphate dehydrogenase *gap1* in the wild-type *S. clavuligerus* doubled clavulanic acid titers most likely because of higher diversion of glyceraldehyde-3-phosphate to clavulanic acid biosynthesis rather than glycolysis pathway.[88] An interesting finding was that polyphosphate kinase *ppk* gene deletion causes antibiotics overproduction in *S. lividans*.[89] *In vivo*, in condition of phosphate limitation, Ppk likely acts as a nucleoside di-phosphate kinase regenerating ATP from ADP and polyphosphates.[90] The absence of this important ATP regenerating enzyme is predicted to lead to the lowering of the intracellular energetic charge (low ATP/ADP ratio), which further triggers a strong activation of central metabolism coupled to respiratory chain in order to regenerate the necessary ATP.[90] The increased antibiotics production can be caused by higher supply of precursors from the central carbon metabolism, though regulatory events due to lower ATP/ADP ratio are probably important as well.

7.5.4 Changing morphology

Filamentous bacteria pose certain challenges in connection with large scale fermentation processes when compared to unicellular organisms. During fermentation there are problems with nutrients and oxygen transfer and shear damage of cells from agitation. Filtration/centrifugation in the downstream processing is also complicated by the high viscosity of the fermentation broth. In many strains hyphae clump together in so-called pellets, which are spherical hyphal clumps. The cells in the center of the pellet may lack nutrients and oxygen, and these cells are often stressed or dead resulting in an inactive core of the pellets. Effectively, it is only the cells on the surface of the pellets that grow and synthesize the product. It has been observed for *S. noursei*, that when pellet size increased above a certain critical value, the growth and nystatin production ceased.[91] In *S. tendae* submerged cultivations it was found that maximal productivity of nikkomycins was achieved at a pellet diameter of 1.4 mm, and the productivity was lower for both smaller and bigger pellets.[92] Possibly, pellet formation protects *S. tendae* mycelia from hyphal fragmentation and hereby has a positive effect until the pellets reach a critical diameter where nutrient diffusion becomes a problem. There is still much unknown about the impact of morphology on production and in some cases formation of larger

pellets seems to be desirable. For instance, *S. lividans* engineered for production of a hybrid antibiotic, was making the product only in the presence of compact mycelial pellets.[93]

Apart from physical methods that can be applied to decrease pellets formation (high agitation rates, addition of viscosity-increasing agents, etc.), one can change the morphology at the genetic level. The SsgA protein that accumulates just before the onset of spore formation was first described in *S. griseus* (Yamamoto 1995). The protein is involved in formation of cross walls called septa. Vegetative septa are thin single-layer walls that divide vegetative mycelium into compartments, in contrast sporulation septa are thick double-layer walls that often form synchronically and divide sporogenic hyphae into chains of spores.[94] When *ssgA* was over-expressed, *S. griseus* failed to sporulate in the submerged culture and the mycelium was highly fragmented.[95] A homologous gene was also found in *S. coelicolor*; and overexpression of this gene caused increased septa formation in the vegetative mycelium.[96] This resulted in a more fragmented mycelium and smaller and fewer pellets. The specific growth rate was increased 40 to 70% depending on the medium and the final concentration of undecylprodigiosin was higher by an order of magnitude.[97] Over-expressing the *S. coelicolor ssgA* gene in other *Streptomyces* had similar effects: in *S. roseosporus* and *S. lividans* mycelial clumps got smaller and *S. venezuelae* did not longer sporulate in liquid medium. Excess of SsgA also improved growth and productivity of a *S. lividans* strain making a model protein tyrosinase, i.e. the specific growth rate was increased by 45% and tyrosinase productivity more than 2 fold.[97]

7.5.5 Improving oxygen supply

Because of the filamentous character of *Streptomyces*, oxygen supply during fermentations is often limited. *Streptomyces* are strictly aerobic organisms and proper oxygen supply is essential for obtaining good biomass yields in the initial growth phase. Sufficient aeration is also necessary for some oxygen-dependent steps in product biosynthesis. For instance, there are two oxygen-dependent reactions in the biosynthesis of erythromycin.[98, 99] Oxygen uptake of *Saccharopolyspora erythraea* was improved by over-expressing the bacterial hemoglobin gene *vhb* from *Vitreoscilla* sp. on an integrating vector, resulting in 60% higher erythromycin production.[100] Positive effect of *vhb* expression on antibiotics production has also been observed in other *Streptomyces* species.[101]

7.5.6 Improving secretion and reducing degradation of recombinant proteins

The commonly used host for recombinant protein production *E. coli* often accumulates large amounts of heterologous proteins as insoluble aggregates in the form of inclusion bodies. Recombinant protein purification consequently has to involve cell disruption, separation of inclusion bodies as well as laborious protein refolding. Production of pharmaceutical proteins in secreted form has some major advantages over intracellular production: the secreted target protein is usually natively folded, yields can be as high as or higher than that obtained from intracellular *E. coli* accumulation and finally there is a reduced requirement for expensive extraction and purification procedures. In *E. coli*, as Gram-negative bacterium, secreted proteins accumulate in the periplasm due to the presence of an outer membrane. Secretion of proteins directly into the fermentation medium can be obtained using Gram-positive bacteria as production host, *i.e. B. subtilis*. Lately *S.*

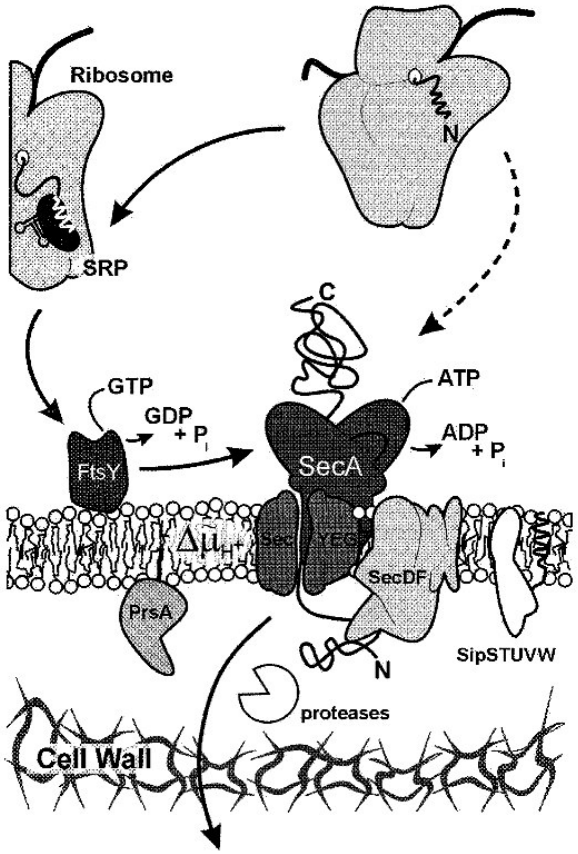
lividans has also been employed for a number of recombinant protein processes.[9, 10] *Streptomyces* are good at secreting proteins because their survival in soil depends on their ability to secrete enzymes for degrading complex substrates. In the genome of *S. coelicolor* 819 genes encoding secreted proteins were predicted. These proteins include various enzymes for the degradation of exogenous substrates: proteases/peptidases, chitinases/chitosanases, cellulases/endoglucanases, amylases and pectate lyases. Extracellular proteins accumulate in the culture broth of *S. coelicolor* mostly during transition and stationary phases.[102]

Proteins in Gram-positive organisms are transported through the Sec and the twin-arginine translocation (Tat) pathways, but some polypeptides can be exported by ABC transporters and for some specific proteins other Sec/Tat-independent mechanisms exist (Paul Dyson, Swansea University, UK, presentation at Genetics of Industrial Microorganisms 2006).

The Sec pathway is the most commonly used pathway. In *B. subtilis* deletion of the gene encoding SecA, a major player in the Sec pathway, reduced the extracellular proteome by 90%, suggesting that all these proteins used Sec transport.[103] The Sec pathway consists of a few components (Figure 7-5). Membrane proteins SecY, SecE and SecG form a channel through which proteins are transported.[104-106] SecA is the main driver of the transport process: it exists in a free form in the cytoplasm and in the SecYEG-bound form. SecA binds pre-proteins and, using the energy from ATP hydrolysis, step by step pushes the largely unfolded pre-proteins through the cytoplasmic membrane channel.[107-110] SecDF associate with the SecYEG complex and contribute to the efficiency of protein secretion.[111-114] Targeting of proteins to the SecYEG channel can occur in a post-translational process by chaperones or via the SRP-mediated pathway in a co-translational fashion. In the so-called co-translational translocation, protein translocation is coupled to their synthesis on the ribosomes.[115, 116] When these proteins are emerging from the ribosome, their signal peptide is recognized by the signal recognition particle (SRP), which binds to the SRP receptor, FtsY protein[117], conserved among all studied bacteria, including *S. lividans*.[118, 119] In *B. subtilis*, SRP is required for the targeting of most secretory proteins.[120] Other proteins are first completely synthesized and then exported in a so called post-translational translocation. It is important that the folding of those proteins in the cytoplasm is prevented, because fully folded proteins can not be transported through the Sec system. A secretion-dedicated *E. coli* chaperone SecB stabilizes pre-proteins inside the cytoplasm, so they preserve the translocation-competent state.[121, 122] SecB binds to SecA and in this way directs the proteins towards the translocation channel.[123, 124] SecB homologs supporting post-translational targeting are not present in Gram-positive organisms.[125] The role of other common chaperones and heat shock proteins (GroE and DnaK series) in the process of protein secretion is not very clear, but there is some evidence that they enhance secretion of proteins by reducing formation of protein agglomerates. Higher expression of GroE and DnaK operons, mediated by deletion of a negative regulator *hrcA*, combined with overexpression of extracytoplasmic molecular chaperone PsrA improved production of single chain antibody fragments in *B. subtilis.*[126, 127] In *Streptomyces* the GroE operon is under negative regulation by the HrcA repressor[128] and the DnaK operon is repressed by HspR.[45, 129, 130] Deletion mutants Δ*hrcA* and Δ*hspR* have been constructed in *S. coelicolor, S. lividans* and *S. albus* strains.[45, 128-130] So far, only the effect of *hrcA* and *hspR* deletion was investigated in *S. lividans* and, here, no positive effect on Sec-

dependent protein secretion was observed (personal communication with Jozef Anné, Katholieke Universiteit Leuven, Belgium).

Figure 7-5. Co-translational (whole line) and post-translational (dashed line) translocation of proteins via Sec pathway.[125]



The twin-arginine translocation (Tat) pathway can transport already folded proteins often bound to redox cofactors. The pathway only uses proton-motive force as the energy source for transport. The proteins transported through the Tat pathway possess a signal peptide with a particular motif S/T-**R-R**-x-Φ-Φ (with Φ being a hydrophobic residue) with two neighboring arginines. The Tat pathway is well studied in *E. coli*[131] and *B. subtilis*[132] and was recently found in *S. lividans*[133]. The search for Tat-dependent signal peptides in the sequenced genomes of *S. coelicolor* and *S. avermitilis* gave a list of 230 proteins, which is the highest number of predicted Tat-substrates for a bacterial genome.[134] Very recently, experimental data provided strong evidence that the Tat system is used as a major general export pathway in *Streptomyces*.[135] *S. lividans* with deletions of principal components of the Tat pathway, such as TatC or TatB, had retarded growth as well as impaired morphological differentiation on solid medium, confirming the importance of Tat-dependent secretion.[134] Some proteins require Tat-dependent secretion for being active. For instance, when xylanase C, a Tat-substrate, was directed to the Sec pathway through replacement of the signal peptide, the result was an inactive protein, which was quickly degraded after secretion.[136] Overexpression of the *tatABC* genes in *S. lividans*, however, improved xylanase C production.[137] To compare the transport through the Sec and Tat pathways, two human proteins (tumor necrosis

factor and interleukin 10) were expressed in *S. lividans* with fused Sec- or Tat-dependent signal peptides. The secretion was lower for the Tat pathway, which might be attributed to suboptimal combinations of signal peptide and target protein or to the lower capacity and speed or higher energetics of the Tat pathway itself.[138] The authors made an interesting discovery: secretion through the Sec pathway was several fold (up to 15) improved in the strain with an impaired Tat-pathway (Δ*tatB*). It seems that the presence of Tat proteins has a negative effect on the Sec pathway, because it has also been shown that in a *S. lividans* strain with overexpressed TatABC proteins, the secretion of Sec-substrates was reduced.[137] In some cases combination of both secretion pathways can be beneficial: xylanase B production was improved in *S. lividans* when xylanase B1 was engineered to be secreted by Sec pathway and xylanase B2 by Tat pathway.[139] The signal peptide, which directs the pre-protein towards a secretion pathway, is an important determinant of how efficiently the protein is secreted. A library of naturally occurring signal peptides from *B. subtilis* has recently been constructed.[140] The library screening was successfully used to optimize production of cutinase from *Fusarium solani pisi* and of a cytoplasmatic esterase. However, the signal peptide, which was optimal for one protein was not efficient for another protein secretion.[140] This study highlights that secretion depends both of the signal peptide sequence and of the sequence of the mature part, hence signal peptides should be optimized individually for different proteins. Li et al. mutagenized the signal peptide of xylanase C, exported through the Tat pathway, but did not achieve improvement of secretion in *S. lividans*.[141] In contrast, it was possible to enhance the secretion of mouse tumor necrosis factor alpha through Sec pathway by decreasing the positive charge in the signal peptide sequence.[142, 143]

During the export through the Sec or Tat pathways the signal peptide has to be cleaved off by signal peptidases. There are four known type I signal peptidases in *S. lividans*: *sipW*, *sipX*, *sipY* and *sipZ.*[144] Deletion of one of the peptidases does not influence protein secretion, with an exception of *sipY* deletion, which decreases extracellular proteins concentration and gives a sporulation delay.[145] Overexpression of *sip* genes had a positive effect on protein processing in *S. lividans* (personal communication with Nick Geukens, Katholieke Universiteit Leuven, Belgium), as also observed in *B. subtilis*.[146, 147] Recently, it was also shown that phage-shock protein A (PspA), which is supposed to play a role in the maintenance of the proton motive force (PMF),[148] affects the protein secretion yield in *S. lividans*. As also observed in *E. coli*,[149] *pspA* overexpression was found to improve the Tat-dependent protein secretion in *S. lividans*. The effect on Sec-dependent secretion was less pronounced and appeared to be protein dependent.[150] During the whole process of translocation as well as after secretion the proteins are targets of endo- and extracellular peptidases/proteases. *S. lividans* is a popular strain for heterologous protein production due to its low protease activity.[151] Growing *S. lividans* in a pelleted form further reduces the activity of extracellular peptidases.[152] Some intra- and extracellular proteases have, however, been identified in *S. lividans*.[153, 154] Actinomycetes were also found to contain 20S proteosome, a structure made of self-compartmentalizing protease, which otherwise occurs only in eukaryotes and archaea.[155] Deletion of 20S proteosome in *S. lividans* had a positive effect on production of two heterologous proteins: soluble human tumour necrosis factor receptor II (shuTNFRII), secreted via Sec-pathway and salmon calcitonin (sCT), secreted via Tat-pathway, but did not affect the production of soluble human tumour

necrosis factor receptor I (shuTNFRI).[156] This suggests that only some of the proteins are subjected to the proteosome activity and can benefit from proteosome deletion.

7.5.7 Changing regulation

By changing the central transcriptional regulators it is possible to introduce global changes in the metabolism. The outcome of such a modification is difficult to predict, but the increasing amounts of protein-DNA interactions and transcriptional data will hopefully allow designing modifications of this magnitude in the future.

One of the best studied global regulators of secondary metabolism is the regulatory protein AfsR. It was first discovered in *S. coelicolor* as an element that induced overproduction of the signaling molecule A-factor and of the antibiotics actinorhodin, undecylprodigiosin[157] and CDA[158] when cloned into *S. lividans*.[158] The effect on antibiotics production was mediated by enhancing transcription of secondary metabolic genes. Antibiotics production was reduced but not eliminated in a Δ*afsR* mutant.[159] A homologue of *asfR* gene was found in *S. peucetius*, and when overexpressed it resulted in antibiotics overproduction in *S. peucetius* (doxorubicin), *S. lividans* TK 24 (γ-actinorhodin), *S. clavuligerus* (clavulanic acid) and *S. griseus* (streptomycin). AfsR is believed to activate transcription of a sigma-like gene *afsR2*, which in turn enhances transcription of the secondary metabolic genes by a yet unknown mechanism.[160] Overexpression of *afsR2* caused antibiotics overproduction in *S. coelicolor*[161], *S. lividans*[162], *S. avermitilis*[163] and *S. noursei*[164].

Another regulator is S-adenosylmethionine (SAM), the methyl donor in methylation of nucleic acids, proteins and small metabolites. SAM levels in *S. coelicolor* peak around the point when actinorhodin appears in the medium.[165] When SAM synthase coded by the *metK* gene was overexpressed in *S. coelicolor*, the intracellular SAM levels increased and it caused overproduction of actinorhodin, prodigiosins and calcium-dependent antibiotic (CDA). Actinorhodin biosynthesis was influenced through increased expression of the actinorhodin cluster activator *actII-4*. The antibiotics were also synthesized earlier than normally. Similar effect was observed when SAM was added into the culture of a wild type strain.[165] Positive effect of SAM on antibiotics production has also been observed in *S. lividans*[166] and other *Streptomyces* spp.[167]

7.5.8 Modifying the product

The increasing resistance of pathogens to antibiotics in current use raises the need for discovery of new antibiotics. The screening of soil isolates continues to bring new antibiotic-producing *Streptomyces*, but the process is slowing down as more and more species become known. It is speculated that 99% of all microorganisms can not be cultivated under laboratory conditions. To capture the antibiotics that could be produced by non-cultivatable *Streptomyces* and other microorganisms, researchers have started to look for DNA instead of microbes themselves. DNA can be isolated from environmental samples and incorporated into libraries, which can for example be screened by hybridization of the conserved genes sequences (like PKS).[168-170] Subsequently the discovered genes can be expressed in a heterologous host to study possible products encoded by the DNA.

Another way of obtaining new products is by varying the genes coding for their biosynthesis. Most of the work has been done on polyketide synthases, which are large modular enzymes, encoding the synthesis of complex polyketide molecules. Polyketides are made in a mechanism that resembles the biosynthesis of long fatty

acids, but the units added at each step are often larger than an acetyl unit and moreover the molecules are subjected to various transformations during extension. Polyketide synthases consist of modules, each module containing transferases, reductases, dehydrases and other enzymes in different amounts and order. The modular structure and modules position in the cluster defines the final product. This unique organization of polyketide synthases allows generation of new products by combinatorial biosynthesis.[74] Native parts of the module can be substituted with analogous enzymes from different strains: replacement of the acyl carrier protein in *S. coelicolor* actinorhodin cluster by homologues from other clusters resulted in production of new actinorhodin analogues.[171] Other strategies are to change the substrate specificity of PKS enzymes by mutations[172] or to shuffle the positions of the modules in the PKS[173] or to delete certain enzyme functionalities[174]. Finally, it has been shown possible to reconstruct a desirable PKS from required elements[175].
Other genes that are very amenable for "genetic molecules design" are non-ribosomal peptide synthases (NRPS), where different subunits are responsible for incorporating a particular amino acid into the peptide. A number of new daptomycin derivatives with antibacterial activity were obtained by substitution of the native NRPS subunit by a homolog from a different cluster.[176] Feeding of slightly changed antibiotics precursors which can still be recognized and processed by the biosynthetic enzymes can also be used as a strategy to obtain new products.[177]

7.6 Perspectives

Over the past 50 years *Streptomyces* have been a rich source for production of many valuable secondary metabolites. With the recent development in genomic techniques *Streptomyces* studies have entered the post-genomic level, and this may allow better understanding of the metabolism, morphological differentiation and overall regulation in these complex bacteria. The accumulated knowledge will allow us to faster optimize industrial *Streptomyces* strains for the production of secondary metabolites, enzymes and other products. We envisage improvement of genetic manipulation techniques, which will facilitate the molecular biology of streptomycetes.
We hope that large amounts of projects on screening of environmental DNA and RNA samples will bear fruits and new bioactive compounds, including efficient antibiotics, will be discovered to help us fight infections caused by drug-resistant pathogenic bacteria and other diseases. Thus, we foresee that *Streptomyces* will continue to play a very important role as workhorses for the production of natural products that serves as important pharmaceuticals also in the next 50 years.

7.7 Acknowledgements

We thank Nick Geukens (Laboratory of Bacteriology, Rega Institute for Medical Research, Katholieke Universiteit Leuven, Belgium), Paul Dyson (Institute of Life Science, School of Medicine, University of Wales Swansea, UK), Marie Jolie Virolle (Laboratoire de Biologie et Genetique Moleculaire de l'Institut de Genetique et Microbiologie, France) and Jette Thykkær (Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, Denmark) for comments on the manuscript.

Table 7-2. Useful links for *Streptomyces*

http://streptomyces.org.uk	*Streptomyces* resource, contains links to *S. coelicolor* A3(2) genome annotation, pathway/genome database ScoCYC, microarray and proteomics data
http://www.surrey.ac.uk/SBMS/Fgenomics/Microarrays/html/Facility.html	Microarray production facility in Surrey University, making *S. coelicolor* DNA arrrays
http://genome-www5.stanford.edu/	Stanford MicroArray Database contains some expression data from *Streptomyces*
Strep-microarray.sbs.surrey.ac.uk/stropE.html	Operon predictions in *S. coelicolor*[47]
http://streptomyces.org.uk/redirect	Web tools for design of REDIRECT knock-out in *S. coelicolor*

7.8 References

1. Schatz, A., Bugie, E., and Waksman, S., Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria, *Proc Soc Exptl Biol Med*, 55, 66, 1944.
2. Tuncer, M., Kuru, A., Isikli, M., Sahin, N., and Celenk, F.G., Optimization of extracellular endoxylanase, endoglucanase and peroxidase production by *Streptomyces* sp. F2621 isolated in Turkey, *J Appl Microbiol*, 97, 783, 2004.
3. Macedo, J.M., Gottschalk, L.M., and Bon, E.P., Lignin peroxidase and protease production by *Streptomyces viridosporus* T7A in the presence of calcium carbonate. Nutritional and regulatory carbon sources, *Appl Biochem Biotechnol*, 77-79, 735, 1999.
4. Yokoyama, K., Nio, N., and Kikuchi, Y., Properties and applications of microbial transglutaminase, *Appl Microbiol Biotechnol*, 64, 447, 2004.
5. Okeke, B.C., Jr, F.W., Use of starch and potato peel waste for perchlorate bioreduction in water, *Sci Total Environ*, 347, 35, 2005.
6. Sette, L.D., Oliveira, V.M., and Manfio, G.P., Isolation and characterization of alachlor-degrading actinomycetes from soil, *Antonie Van Leeuwenhoek*, 87, 81, 2005.
7. Okeke, B.C., Frankenberger, W.T., Biodegradation of methyl tertiary butyl ether (MTBE) by a bacterial enrichment consortia and its monoculture isolates, *Microbiol Res*, 158, 99, 2003.
8. Sariaslani, F.S., Trower, M.K., and Omer, C.A., Constitutive expression of P450SOY and ferredoxin-SOY in *Streptomyces*, and biotransformation of chemicals by recombinant organisms, patent US 9210885, 1993.
9. Garvin, R.T., Malek, L.T., An expression system for the secretion of bioactive human granulocyte macrophage colony stimulating factor (GM-CSF) and other heterologous proteins from *Streptomyces*, patent 89113607.9, 1989.
10. DeSanti, C.L., Strohl, W.R., Soluble recombinant endostatin and method of making same from *Streptomyces* sp, patent US 0009747, 2000.
11. Kieser T, Bibb M, Buttner M, Chater K, D.A. H. Practical *Streptomyces* Genetics, The John Innes Foundation, Norwich, 2000.
12. Hopwood, D.A., *Streptomyces* genes: from Waksman to Sanger, *J Ind Microbiol Biotechnol*, 30, 468, 2003.
13. Sianidis, G., Pozidis, C., Becker, F., Vrancken, K., Sjoeholm, C., Karamanou, S., Takamiya-Wik, M., L, v.M., Schaefer, T., Anne, J., and Economou, A.,

Functional large-scale production of a novel *Jonesia* sp. xyloglucanase by heterologous secretion from *Streptomyces lividans*, *J Biotechnol*, 2005.
14. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., and Goble, A., Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature*, 417, 141, 2002.
15. Lin, Y.S., Kieser, H.M., Hopwood, D.A., and Chen, C.W., The chromosomal DNA of *Streptomyces lividans* 66 is linear, *Mol Microbiol*, 141103, 1994.
16. Lezhava, A., Mizukami, T., Kajitani, T., Kameoka, D., Redenbach, M., Shinkawa, H., Nimi, O., and Kinashi, H., Physical map of the linear chromosome of *Streptomyces griseus*, *J Bacteriol*, 177, 6492, 1995.
17. Chen, C.W., Huang, C.H., Lee, H.H., Tsai, H.H., and Kirby, R., Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes, *Trends Genet*, 18, 522, 2002.
18. Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., and Omura, S., Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*, *Nat Biotechnol*, 21, 526, 2003.
19. MacNeil, D.J., Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*, *J Bacteriol*, 170, 5607, 1988.
20. Mazodier, P., Petter, R., and Thompson, C., Intergeneric conjugation between *Escherichia coli* and *Streptomyces species*, *J Bacteriol*, 171, 3583, 1989.
21. Flett, F., Mersinias, V., and Smith, C.P., High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes, *FEMS Microbiol Lett*, 155, 223, 1997.
22. Boccard, F., Smokvina, T., Pernodet, J.L., Friedmann, A., and Guerineau, M., The integrated conjugative plasmid pSAM2 of *Streptomyces ambofaciens* is related to temperate bacteriophages, *EMBO J*, 8, 973, 1989.
23. Boccard, F., Smokvina, T., Pernodet, J.L., Friedmann, A., and Guerineau, M., Structural analysis of loci involved in pSAM2 site-specific integration in *Streptomyces* , *Plasmid*, 21, 59, 1989.
24. Combes, P., Till, R., Bee, S., and Smith, M.C., The *Streptomyces* genome contains multiple pseudo-attB sites for the (phi)C31-encoded site-specific recombination system, *J Bacteriol*, 184, 5746, 2002.
25. Murakami, T., Holt, T.G., and Thompson, C.J., Thiostrepton-induced gene expression in *Streptomyces lividans*, *J Bacteriol*, 171, 1459, 1989.
26. DeSanti, C.L., Strohl, W.R., Characterization of the *Streptomyces* sp. strain C5 *snp* locus and development of *snp*-derived expression vectors, *Appl Environ Microbiol*, 69, 1647, 2003.
27. Nikodinovic, J., Priestley, N.D., A second generation *snp*-derived *Escherichia coli* - *Streptomyces* shuttle expression vector that is generally transferable by conjugation, *Plasmid*, 2006.
28. Jensen, P.R., Hammer, K., The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters, *Appl Environ Microbiol*, 64, 82, 1998.
29. Hammer, K., Mijakovic, I., and Jensen, P.R., Synthetic promoter libraries - tuning of gene expression, *Trends Biotechnol*, 24, 53, 2006.

30. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C., A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*, *Nucleic Acids Res*, 21, 3329, 1993.
31. Datsenko, K.A., Wanner, B.L., One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products, *Proc Natl Acad Sci U S A*, 97, 6640, 2000.
32. Gust, B., Challis, G.L., Fowler, K., Kieser, T., and Chater, K.F., PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin, *Proc Natl Acad Sci U S A*, 100, 1541, 2003.
33. Khodakaramian, G., Lissenden, S., Gust, B., Moir, L., Hoskisson, P.A., Chater, K.F., and Smith, M.C., Expression of Cre recombinase during transient phage infection permits efficient marker removal in *Streptomyces*, *Nucleic Acids Res*, 34, 2006.
34. Shin, S.J., Wu, C.W., Steinberg, H., and Talaat, A.M., Identification of novel virulence determinants in *Mycobacterium paratuberculosis* by screening a library of insertional mutants, *Infect Immun*, 74, 3825, 2006.
35. Pobigaylo, N., Wetter, D., Szymczak, S., Schiller, U., Kurtz, S., Meyer, F., Nattkemper, T.W., and Becker, A., Construction of a large signature-tagged mini-Tn5 transposon library and its application to mutagenesis of *Sinorhizobium meliloti*, *Appl Environ Microbiol*, 72, 4329, 2006.
36. Kumar, A., Seringhaus, M., Biery, M.C., Sarnovsky, R.J., Umansky, L., Piccirillo, S., Heidtman, M., Cheung, K.H., Dobry, C.J., Gerstein, M.B., Craig, N.L., and Snyder, M., Large-scale mutagenesis of the yeast genome using a Tn7-derived multipurpose transposon, *Genome Res*, 14, 1975, 2004.
37. Wallace, L.J., Ward, J.M., and Richmond, M.H., The *tnpR* gene product of TnA is required for transposition immunity, *Mol Gen Genet*, 184, 87, 1981.
38. Herron, P.R., Hughes, G., Chandra, G., Fielding, S., and Dyson, P.J., Transposon Express, a software application to report the identity of insertions obtained by comprehensive transposon mutagenesis of sequenced genomes: analysis of the preference for *in vitro* Tn5 transposition into GC-rich DNA, *Nucleic Acids Res*, 32, 2004.
39. Volff, J.N., Altenbuchner, J., Genetic instability of the *Streptomyces* chromosome, *Mol Microbiol*, 27, 239, 1998.
40. Bishop, A., Fielding, S., Dyson, P., and Herron, P., Systematic insertional mutagenesis of a streptomycete genome: a link between osmoadaptation and antibiotic production, *Genome Res*, 14, 893, 2004.
41. Chakraburtty, R., Bibb, M., The ppGpp synthetase gene (*relA*) of *Streptomyces coelicolor* A3(2) plays a conditional role in antibiotic production and morphological differentiation, *J Bacteriol*, 179, 5854, 1997.
42. Chakraburtty, R., White, J., Takano, E., and Bibb, M., Cloning, characterization and disruption of a (p)ppGpp synthetase gene (*relA*) of *Streptomyces coelicolor* A3(2), *Mol Microbiol*, 19, 357, 1996.
43. Huang, J., Lih, C.J., Pan, K.H., and Cohen, S.N., Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays, *Genes Dev*, 15, 3183, 2001.

44. Karoonuthaisiri, N., Weaver, D., Huang, J., Cohen, S.N., and Kao, C.M., Regional organization of gene expression in *Streptomyces coelicolor*, *Gene*, 353, 53, 2005.
45. Bucca, G., Brassington, A.M., Hotchkiss, G., Mersinias, V., and Smith, C.P., Negative feedback regulation of *dnaK* , *clpB* and *lon* expression by the DnaK chaperone machine in *Streptomyces coelicolor* , identified by transcriptome and *in vivo* DnaK-depletion analysis, *Mol Microbiol*, 50, 153, 2003.
46. Mehra, S., Lian, W., Jayapal, K.P., Charaniya, S.P., Sherman, D.H., and Hu, W.S., A framework to analyze multiple time series data: a case study with *Streptomyces coelicolor*, *J Ind Microbiol Biotechnol*, 33, 159, 2006.
47. Laing, E., Mersinias, V., Smith, C.P., and Hubbard, S.J., Analysis of gene expression in operons of *Streptomyces coelicolor*, *Genome Biol*, 7, 2006.
48. McArthur, M., Bibb, M., *In vivo* DNase I sensitivity of the *Streptomyces coelicolor* chromosome correlates with gene expression: implications for bacterial chromosome structure, *Nucleic Acids Res*, 2006.
49. Lum, A.M., Huang, J., Hutchinson, C.R., and Kao, C.M., Reverse engineering of industrial pharmaceutical-producing actinomycete strains using DNA microarrays, *Metab Eng*, 6, 186, 2004.
50. Hesketh, A.R., Chandra, G., Shaw, A.D., Rowland, J.J., Kell, D.B., Bibb, M.J., and Chater, K.F., Primary and secondary metabolism, and post-translational protein modifications, as portrayed by proteomic analysis of *Streptomyces coelicolor*, *Mol Microbiol*, 46, 917, 2002.
51. Wittmann, C., Kromer, J.O., Kiefer, P., Binz, T., and Heinzle, E., Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria, *Anal Biochem*, 327, 135, 2004.
52. Borodina I, Siebring J, Zhang J, Smith C, Van Keulen G, Dijkhuizen L, Nielsen J. Antibiotics overproduction in *Streptomyces coelicolor* A3(2) mediated by phosphofructokinase A2 deletion, manuscript in preparation, 2006.
53. Kromer, J.O., Sorgenfrei, O., Klopprogge, K., Heinzle, E., and Wittmann, C., In-depth profiling of lysine-producing *Corynebacterium glutamicum* by combined analysis of the transcriptome, metabolome, and fluxome, *J Bacteriol*, 186, 1769, 2004.
54. Avignone Rossa, White, J., Kuiper, A., Postma, P.W., Bibb, M., and Teixeira de Mattos, M.J., Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*, *Metab Eng*, 4, 138, 2002.
55. Kim, H.B., Smith, C.P., Micklefield, J., and Mavituna, F., Metabolic flux analysis for calcium dependent antibiotic (CDA) production in *Streptomyces coelicolor*, *Metab Eng*, 6, 313, 2004.
56. Naeimpoor, F., Mavituna, F., Metabolic flux analysis in *Streptomyces coelicolor* under various nutrient limitations, *Metab Eng*, 2, 140, 2000.
57. Gunnarsson, N., Mortensen, U.H., Sosio, M., and Nielsen, J., Identification of the Entner-Doudoroff pathway in an antibiotic-producing actinomycete species, *Mol Microbiol*, 52, 895, 2004.
58. Borodina, I., Scholler, C., Eliasson, A., and Nielsen, J., Metabolic network analysis of *Streptomyces tenebrarius* , a *Streptomyces* species with an active entner-doudoroff pathway, *Appl Environ Microbiol*, 71, 2294, 2005.
59. Inbar, L., Lapidot, A., 13C nuclear magnetic resonance and gas chromatography-mass spectrometry studies of carbon metabolism in the actinomycin D

producer *Streptomyces parvulus* by use of 13C-labeled precursors, *J Bacteriol*, 173, 7790, 1991.
60. Jonsbu, E., Christensen, B., and Nielsen, J., Changes of *in vivo* fluxes through central metabolic pathways during the production of nystatin by *Streptomyces noursei* in batch culture, *Appl Microbiol Biotechnol*, 56, 93, 2001.
61. Gunnarsson, N., Bruheim, P., and Nielsen, J., Glucose metabolism in the antibiotic producing actinomycete *Nonomuraea* sp. ATCC 39727, *Biotechnol Bioeng*, 88, 652, 2004.
62. Borodina, I., Krabben, P., and Nielsen, J., Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism, *Genome Res*, 15, 820, 2005.
63. Borodina, I., Nielsen, J., From genomes to *in silico* cells via metabolic networks, *Curr Opin Biotechnol*, 16, 350, 2005.
64. Pharkya, P., Maranas, C.D., An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems, *Metab Eng*, 8, 1, 2006.
65. Patil, K.R., Rocha, I., Forster, J., and Nielsen, J., Evolutionary programming as a platform for *in silico* metabolic engineering, *BMC Bioinformatics*, 6308, 2005.
66. Patil, K.R., Nielsen, J., Uncovering transcriptional regulation of metabolism by using metabolic network topology, *Proc Natl Acad Sci U S A*, 102, 2685, 2005.
67. Alper, H., Jin, Y.S., Moxley, J.F., and Stephanopoulos, G., Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*, *Metab Eng*, 7, 155, 2005.
68. Alper, H., Miyaoku, K., and Stephanopoulos, G., Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets, *Nat Biotechnol*, 23, 612, 2005.
69. Rodriguez, E., Hu, Z., Ou, S., Volchegursky, Y., Hutchinson, C.R., and McDaniel, R., Rapid engineering of polyketide overproduction by gene transfer to industrially optimized strains, *J Ind Microbiol Biotechnol*, 30, 480, 2003.
70. Rowe, C.J., Cortes, J., Gaisser, S., Staunton, J., and Leadlay, P.F., Construction of new vectors for high-level expression in actinomycetes, *Gene*, 216, 215, 1998.
71. Kao, C.M., Katz, L., and Khosla, C., Engineered biosynthesis of a complete macrolactone in a heterologous host, *Science*, 265, 509, 1994.
72. Ziermann, R., Betlach, M.C., Recombinant polyketide synthesis in *Streptomyces*: engineering of improved host strains, *Biotechniques*, 26, 106, 1999.
73. Zirkle, R., Ligon, J.M., and Molnar, I., Heterologous production of the antifungal polyketide antibiotic soraphen A of *Sorangium cellulosum* So ce26 in *Streptomyces lividans*, *Microbiology*, 150, 2761, 2004.
74. McDaniel, R., Ebert-Khosla, S., Hopwood, D.A., and Khosla, C., Engineered biosynthesis of novel polyketides, *Science*, 262, 1546, 1993.
75. Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C., and Julien, B., Cloning and heterologous expression of the epothilone gene cluster, *Science*, 287, 640, 2000.
76. Yanai, K., Murakami, T., and Bibb, M., Amplification of the entire kanamycin biosynthetic gene cluster during empirical strain improvement of *Streptomyces kanamyceticus*, *Proc Natl Acad Sci U S A*, 2006.
77. Bejar, S., Belghith, K., and Ellouz, R., Glucose isomerase of *S. violaceoniger* . Fundamental and applied aspects, *Arch Inst Pasteur Tunis*, 71, 407, 1994.

78. Gramajo, H.C., Takano, E., and Bibb, M.J., Stationary-phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated, *Mol Microbiol*, 7, 837, 1993.
79. Takano, E., Gramajo, H.C., Strauch, E., Andres, N., White, J., and Bibb, M.J., Transcriptional regulation of the *redD* transcriptional activator gene accounts for growth-phase-dependent production of the antibiotic undecylprodigiosin in *Streptomyces coelicolor* A3(2), *Mol Microbiol*, 6, 2797, 1992.
80. Metsa-Ketela, M., Ylihonko, K., and Mantsala, P., Partial activation of a silent angucycline-type gene cluster from a rubromycin beta producing *Streptomyces* sp. PGA64, *J Antibiot*, 57, 502, 2004.
81. Baltz, R.H., McHenney, M.A., Cantwell, C.A., Queener, S.W., and Solenberg, P.J., Applications of transposition mutagenesis in antibiotic producing streptomycetes, *Antonie Van Leeuwenhoek*, 71, 179, 1997.
82. Lombo, F., Pfeifer, B., Leaf, T., Ou, S., Kim, Y.S., Cane, D.E., Licari, P., and Khosla, C., Enhancing the atom economy of polyketide biosynthetic processes through metabolic engineering, *Biotechnol Prog*, 17, 612, 2001.
83. Ryu, Y.G., Butler, M.J., Chater, K.F., and Lee, K.J., Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*, *Appl Environ Microbiol*, 2006.
84. Li, Y., Ling, H., Li, W., and Tan, H., Improvement of nikkomycin production by enhanced copy of *sanU* and *sanV* in *Streptomyces ansochromogenes* and characterization of a novel glutamate mutase encoded by *sanU* and *sanV*, *Metab Eng*, 7, 165, 2005.
85. McIntyre, J., Bull, A., and Bunch, A., Vancomycin production in batch and continuous culture, *Biotechnol Bioeng*, 49, 412, 1996.
86. Rius, N., Maeda, K., and Demain, A.L., Induction of L-lysine epsilon-aminotransferase by L-lysine in *Streptomyces clavuligerus*, producer of cephalosporins, *FEMS Microbiol Lett*, 144, 207, 1996.
87. Paradkar, A.S., Mosher, R.H., Anders, C., Griffin, A., Griffin, J., Hughes, C., Greaves, P., Barton, B., and Jensen, S.E., Applications of gene replacement technology to *Streptomyces clavuligerus* strain development for clavulanic acid production, *Appl Environ Microbiol*, 67, 2292, 2001.
88. Li, R., Townsend, C.A., Rational strain improvement for enhanced clavulanic acid production by genetic engineering of the glycolytic pathway in *Streptomyces clavuligerus*, *Metab Eng*, 8, 240, 2006.
89. Chouayekh, H., Virolle, M.J., The polyphosphate kinase plays a negative role in the control of antibiotic production in *Streptomyces lividans*, *Mol Microbiol*, 43, 919, 2002.
90. Ghorbel, S., Smirnov, A., Chouayekh, H., Sperandio, B., Esnault, C., Kormanec, J., and Virolle, M.J., Regulation of *ppk* expression and *in vivo* function of Ppk in *Streptomyces lividans* TK24, *J Bacteriol*, 188, 6269, 2006.
91. Jonsbu, E., McIntyre, M., and Nielsen, J., The influence of carbon sources and morphology on nystatin production by *Streptomyces noursei*, *J Biotechnol*, 95, 133, 2002.
92. Vecht-Lifshitz, S.E., Sasson, Y., and Braun, S., Nikkomycin production in pellets of *Streptomyces tendae*, *J Appl Bacteriol*, 72, 195, 1992.
93. Sarra, M., Perez-Pons, J.A., Godia, F., and C, C.A., Importance of growth form on production of hybrid antibiotic by *Streptomyces lividans* TK21 by fed-batch and continuous fermentation, *Appl Biochem Biotechnol*, 75, 235, 1998.

94. Kwak, J., Dharmatilake, A.J., Jiang, H., and Kendrick, K.E., Differential regulation of *ftsZ* transcription during septation of *Streptomyces griseus*, *J Bacteriol*, 183, 5092, 2001.
95. Kawamoto, S., Watanabe, H., Hesketh, A., Ensign, J.C., and Ochi, K., Expression analysis of the *ssgA* gene product, associated with sporulation and cell division in *Streptomyces griseus*, *Microbiology*, 143 (Pt 4), 1077, 1997.
96. van Wezel, G., J, v.d.M., Kawamoto, S., Luiten, R.G., Koerten, H.K., and Kraal, B., *ssgA* is essential for sporulation of *Streptomyces coelicolor* A3(2) and affects hyphal development by stimulating septum formation, *J Bacteriol*, 182, 5653, 2000.
97. G.P., v.W., Krabben, P., Traag, B.A., Keijser, B.J., Kerste, R., Vijgenboom, E., Heijnen, J.J., and Kraal, B., Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological engineering, *Appl Environ Microbiol*, 72, 5283, 2006.
98. Katz, L., Donadio, S., Macrolides, *Biotechnology (Reading, Mass.)*, 28, 385, 1995.
99. Stassi, D., Donadio, S., Staver, M.J., and Katz, L., Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis, *J Bacteriol*, 175, 182, 1993.
100. Brunker, P., Minas, W., Kallio, P.T., and Bailey, J.E., Genetic engineering of an industrial strain of *Saccharopolyspora erythraea* for stable expression of the *Vitreoscilla* haemoglobin gene (*vhb*), *Microbiology*, 144, 2441, 1998.
101. Tabakov, V., Emel'ianova, L.K., Antonova, S.V., and Voeikova, T.A., Effect of the gene for bacterial hemoglobin *vhb* on the effectiveness of the process of *Escherichia coli* - *Streptomyces* interspecies conjugation and production of antibiotics in streptomycetes, *Genetika*, 37, 422, 2001.
102. Kim, D.W., Chater, K., Lee, K.J., and Hesketh, A., Changes in the extracellular proteome caused by the absence of the *bldA* gene product, a developmentally significant tRNA, reveal a new target for the pleiotropic regulator AdpA in *Streptomyces coelicolor*, *J Bacteriol*, 187, 2957, 2005.
103. Hirose, I., Sano, K., Shioda, I., Kumano, M., Nakamura, K., and Yamane, K., Proteome analysis of *Bacillus subtilis* extracellular proteins: a two-dimensional protein electrophoretic study, *Microbiology*, 146, 65, 2000.
104. Brundage, L., Hendrick, J.P., Schiebel, E., Driessen, A.J., and Wickner, W., The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation, *Cell*, 62, 649, 1990.
105. Akimaru, J., Matsuyama, S., Tokuda, H., and Mizushima, S., Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from *Escherichia coli*, *Proc Natl Acad Sci U S A*, 88, 6545, 1991.
106. Manting, E.H., C, v.D.D., Remigy, H., Engel, A., and Driessen, A.J., SecYEG assembles into a tetramer to form the active protein translocation channel, *EMBO J*, 19, 852, 2000.
107. Schiebel, E., Driessen, A.J., Hartl, F.U., and Wickner, W., Delta mu H+ and ATP function at different steps of the catalytic cycle of preprotein translocase, *Cell*, 64, 927, 1991.
108. J, v.d.W., Klose, M., Breukink, E., Demel, R.A., B, d.K., Freudl, R., and Driessen, A.J., Characterization of a *Bacillus subtilis* SecA mutant protein deficient in translocation ATPase and release from the membrane, *Mol Microbiol*, 8, 31, 1993.

109. Papanikou, E., Karamanou, S., Baud, C., Sianidis, G., Frank, M., and Economou, A., Helicase Motif III in SecA is essential for coupling preprotein binding to translocation ATPase, *EMBO Rep*, 5, 807, 2004.
110. Karamanou, S., Vrontou, E., Sianidis, G., Baud, C., Roos, T., Kuhn, A., Politou, A.S., and Economou, A., A molecular switch in SecA protein couples ATP hydrolysis to protein translocation, *Mol Microbiol*, 34, 1133, 1999.
111. Gardel, C., Johnson, K., Jacq, A., and Beckwith, J., The *secD* locus of *E.coli* codes for two membrane proteins required for protein export, *EMBO J*, 9, 3209, 1990.
112. Arkowitz, R.A., Wickner, W., SecD and SecF are required for the proton electrochemical gradient stimulation of preprotein translocation, *EMBO J*, 13, 954, 1994.
113. Economou, A., Pogliano, J.A., Beckwith, J., Oliver, D.B., and Wickner, W., SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF, *Cell*, 83, 1171, 1995.
114. Duong, F., Wickner, W., Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme, *EMBO J*, 16, 2756, 1997.
115. Luirink, J., CM, t.H., CC, v.d.W., Oudega, B., High, S., Dobberstein, B., and Kusters, R., An alternative protein targeting pathway in *Escherichia coli*: studies on the role of FtsY, *EMBO J*, 13, 2289, 1994.
116. Powers, T., Walter, P., Co-translational protein targeting catalyzed by the *Escherichia coli* signal recognition particle and its receptor, *EMBO J*, 16, 4880, 1997.
117. Seluanov, A., Bibi, E., FtsY, the prokaryotic signal recognition particle receptor homologue, is essential for biogenesis of membrane proteins, *J Biol Chem*, 272, 2053, 1997.
118. Palacin, A., R, d.l.F., Valle, I., Rivas, L.A., and Mellado, R.P., *Streptomyces lividans* contains a minimal functional signal recognition particle that is involved in protein secretion, *Microbiology*, 149, 2435, 2003.
119. Palomino, C., Mellado, R.P., The *Streptomyces lividans* cytoplasmic signal recognition particle receptor FtsY is involved in protein secretion, *J Mol Microbiol Biotechnol*, 9, 57, 2005.
120. Yamane, K., Bunai, K., and Kakeshita, H., Protein traffic for secretion and related machinery of *Bacillus subtilis*, *Biosci Biotechnol Biochem*, 68, 2007, 2004.
121. Kumamoto, C.A., Beckwith, J., Mutations in a new gene, *secB*, cause defective protein localization in *Escherichia coli*, *J Bacteriol*, 154, 253, 1983.
122. Baars, L., Ytterberg, A.J., Drew, D., Wagner, S., Thilo, C., van Wijk, K., and de Gier, J., Defining the role of the *Escherichia coli* chaperone SecB using comparative proteomics, *J Biol Chem*, 281, 10024, 2006.
123. Trun, N.J., Silhavy, T.J., The genetics of protein targeting in *Escherichia coli* K12, *J Cell Sci Suppl*, 11, 13, 1989.
124. Crane, J.M., Mao, C., Lilly, A.A., Smith, V.F., Suo, Y., Hubbell, W.L., and Randall, L.L., Mapping of the docking of SecA onto the chaperone SecB by site-directed spin labeling: insight into the mechanism of ligand transfer during protein export, *J Mol Biol*, 353, 295, 2005.

125. van Wely, K.H., Swaving, J., Freudl, R., and Driessen, A.J., Translocation of proteins across the cell envelope of Gram-positive bacteria, *FEMS Microbiol Rev*, 25, 437, 2001.
126. Wu, S.C., Yeung, J.C., Duan, Y., Ye, R., Szarka, S.J., Habibi, H.R., and Wong, S.L., Functional production and characterization of a fibrin-specific single-chain antibody fragment from *Bacillus subtilis*: effects of molecular chaperones and a wall-bound protease on antibody fragment production, *Appl Environ Microbiol*, 68, 3261, 2002.
127. Wu, S.C., Ye, R., Wu, X.C., Ng, S.C., and Wong, S.L., Enhanced secretory production of a single-chain antibody fragment from *Bacillus subtilis* by coproduction of molecular chaperones, *J Bacteriol*, 180, 2830, 1998.
128. Grandvalet, C., Rapoport, G., and Mazodier, P., *hrcA*, encoding the repressor of the *groEL* genes in *Streptomyces albus* G, is associated with a second *dnaJ* gene, *J Bacteriol*, 180, 5129, 1998.
129. Bucca, G., Hindle, Z., and Smith, C.P., Regulation of the *dnaK* operon of *Streptomyces coelicolor* A3(2) is governed by HspR, an autoregulatory repressor protein, *J Bacteriol*, 179, 5999, 1997.
130. Grandvalet, C., Servant, P., and Mazodier, P., Disruption of *hspR*, the repressor gene of the *dnaK* operon in *Streptomyces albus* G, *Mol Microbiol*, 23, 77, 1997.
131. Palmer, T., Sargent, F., and Berks, B.C., Export of complex cofactor-containing proteins by the bacterial Tat pathway, *Trends Microbiol*, 13, 175, 2005.
132. Jongbloed, J.D., Martin, U., Antelmann, H., Hecker, M., Tjalsma, H., Venema, G., Bron, S., JM, v.D., and Muller, J., TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway, *J Biol Chem*, 275, 41350, 2000.
133. Schaerlaekens, K., Schierova, M., Lammertyn, E., Geukens, N., Anne, J., and L, V.M., Twin-arginine translocation pathway in *Streptomyces lividans*, *J Bacteriol*, 183, 6727, 2001.
134. Schaerlaekens, K., L, V.M., Lammertyn, E., Geukens, N., and Anne, J., The importance of the Tat-dependent protein secretion pathway in *Streptomyces* as revealed by phenotypic changes in *tat* deletion mutants and genome analysis, *Microbiology*, 150, 21, 2004.
135. Widdick, D.A., Dilks, K., Chandra, G., Bottrill, A., Naldrett, M., Pohlschroder, M., and Palmer, T., The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*, *Proc Natl Acad Sci U S A*, 2006.
136. Faury, D., Saidane, S., Li, H., and Morosoli, R., Secretion of active xylanase C from *Streptomyces lividans* is exclusively mediated by the Tat protein export system, *Biochim Biophys Acta*, 1699, 155, 2004.
137. De Keersmaeker, S., Vrancken, K., L, V.M., Lammertyn, E., Anne, J., and Geukens, N., Evaluation of TatABC overproduction on Tat- and Sec-dependent protein secretion in *Streptomyces lividans*, *Arch Microbiol*, 2006.
138. Schaerlaekens, K., Lammertyn, E., Geukens, N., S, D.K., Anne, J., and L, V.M., Comparison of the Sec and Tat secretion pathways for heterologous protein production by *Streptomyces lividans*, *J Biotechnol*, 112, 279, 2004.
139. Gauthier, C., Li, H., and Morosoli, R., Increase in xylanase production by *Streptomyces lividans*through simultaneous use of the Sec- and Tat-dependent protein export systems, *Appl Environ Microbiol*, 71, 3085, 2005.

140. Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T., and Eggert, T., Systematic screening of all signal peptides from *Bacillus subtilis* : a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria, *J Mol Biol*, 362, 393, 2006.
141. Li, H., Faury, D., and Morosoli, R., Impact of amino acid changes in the signal peptide on the secretion of the Tat-dependent xylanase C from *Streptomyces lividans*, *FEMS Microbiol Lett*, 255, 268, 2006.
142. Lammertyn, E., L, V.M., Schacht, S., Dillen, C., Sablon, E., A, V.B., and Anne, J., Evaluation of a novel subtilisin inhibitor gene and mutant derivatives for the expression and secretion of mouse tumor necrosis factor alpha by *Streptomyces lividans*, *Appl Environ Microbiol*, 63, 1808, 1997.
143. Lammertyn, E., Desmyter, S., Schacht, S., L, V.M., and Anne, J., Influence of charge variation in the *Streptomyces venezuelae* alpha-amylase signal peptide on heterologous protein production by *Streptomyces lividans*, *Appl Microbiol Biotechnol*, 49, 424, 1998.
144. van Roosmalen, M., Geukens, N., Jongbloed, J.D., Tjalsma, H., Dubois, J.Y., Bron, S., van Dijl, J., and Anne, J., Type I signal peptidases of Gram-positive bacteria, *Biochim Biophys Acta*, 1694, 279, 2004.
145. Palacin, A., Parro, V., Geukens, N., Anne, J., and Mellado, R.P., SipY is the *Streptomyces lividans* type I signal peptidase exerting a major effect on protein secretion, *J Bacteriol*, 184, 4875, 2002.
146. Bolhuis, A., Tjalsma, H., Smith, H.E., de Jong, A., Meima, R., Venema, G., Bron, S., and van Dijl, J., Evaluation of bottlenecks in the late stages of protein secretion in *Bacillus subtilis* , *Appl Environ Microbiol*, 65, 2934, 1999.
147. van Dijl, J., de Jong, A., Vehmaanpera, J., Venema, G., and Bron, S., Signal peptidase I of *Bacillus subtilis*: patterns of conserved amino acids in prokaryotic and eukaryotic type I signal peptidases, *EMBO J*, 11, 2819, 1992.
148. Darwin, A.J., The phage-shock-protein response, *Mol Microbiol*, 57, 621, 2005.
149. DeLisa, M.P., Lee, P., Palmer, T., and Georgiou, G., Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway, *J Bacteriol*, 186, 366, 2004.
150. Vrancken, K., De Keersmaeker, S., Geukens, N., Lammertyn, E., Anne, J., and Van Mellaert, L., *pspA* overexpression in *Streptomyces lividans* improves both Sec- and Tat-dependent protein secretion, *Appl Microbiol Biotechnol*, 2006.
151. Binnie, C., Cossar, J.D., and Stewart, D.I., Heterologous biopharmaceutical protein expression in *Streptomyces*, *Trends Biotechnol*, 15, 315, 1997.
152. Yun, S.I., Yahya, A.R., Malten, M., Cossar, D., Anderson, W.A., Scharer, J.M., and Moo-Young, M., Peptidases affecting recombinant protein production by *Streptomyces lividans*, *Can J Microbiol*, 47, 1137, 2001.
153. Binnie, C., Butler, M.J., Aphale, J.S., Bourgault, R., DiZonno, M.A., Krygsman, P., Liao, L., Walczyk, E., and Malek, L.T., Isolation and characterization of two genes encoding proteases associated with the mycelium of *Streptomyces lividans* 66, *J Bacteriol*, 177, 6033, 1995.
154. Krieger, T.J., Bartfeld, D., Jenish, D.L., and Hadary, D., Purification and characterization of a novel tripeptidyl aminopeptidase from *Streptomyces lividans 66*, *FEBS Lett*, 352, 385, 1994.
155. De Mot, R., Nagy, I., Walz, J., and Baumeister, W., Proteasomes and other self-compartmentalizing proteases in prokaryotes, *Trends Microbiol*, 7, 88, 1999.

156. Hong, B., Wang, L., Lammertyn, E., Geukens, N., Van Mellaert, L., Li, Y., and Anne, J., Inactivation of the 20S in *Streptomyces lividans* and its influence on the production of heterologous proteins, *Microbiology*, 151, 3137, 2005.
157. Horinouchi, S., Hara, O., and Beppu, T., Cloning of a pleiotropic gene that positively controls biosynthesis of A-factor, actinorhodin, and prodigiosin in *Streptomyces coelicolor* A3(2) and *Streptomyces lividans*, *J Bacteriol*, 155, 1238, 1983.
158. Horinouchi, S., Malpartida, F., Hopwood, D.A., and Beppu, T., *afsB* stimulates transcription of the actinorhodin biosynthetic pathway in *Streptomyces coelicolor* A3(2) and *Streptomyces lividans*, *Mol Gen Genet*, 215, 355, 1989.
159. Floriano, B., Bibb, M., *afsR* is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2), *Mol Microbiol*, 21, 385, 1996.
160. Lee, P.C., Umeyama, T., and Horinouchi, S., *afsS* is a target of AfsR, a transcriptional factor with ATPase activity that globally controls secondary metabolism in *Streptomyces coelicolor* A3(2), *Mol Microbiol*, 43, 1413, 2002.
161. Matsumoto, A., Ishizuka, H., Beppu, T., and Horinouchi, S., Involvement of a small ORF downstream of the *afsR* gene in the regulation of secondary metabolism in *Streptomyces coelicolor* A3(2), *Actinomycetologica*, 9, 37, 1995.
162. Vogtli, M., Chang, P.C., and Cohen, S.N., *afsR2*: a previously undetected gene encoding a 63-amino-acid protein that stimulates antibiotic production in *Streptomyces lividans*, *Mol Microbiol*, 14, 643, 1994.
163. Lee, J., Hwang, Y., Kim, S., Kim, E., and Choi, C., Effect of a global regulatory gene, *afsR2*, from *Streptomyces lividans* on avermectin production in *Streptomyces avermitilis*, *J Biosci Bioeng*, 89, 606, 2000.
164. Sekurova, O., Sletta, H., Ellingsen, T.E., Valla, S., and Zotchev, S., Molecular cloning and analysis of a pleiotropic regulatory gene locus from the nystatin producer *Streptomyces noursei* ATCC11455, *FEMS Microbiol Lett*, 177, 297, 1999.
165. Okamoto, S., Lezhava, A., Hosaka, T., Okamoto-Hosoya, Y., and Ochi, K., Enhanced expression of S-adenosylmethionine synthetase causes overproduction of actinorhodin in *Streptomyces coelicolor* A3(2), *J Bacteriol*, 185, 601, 2003.
166. Kim, D.J., Huh, J.H., Yang, Y.Y., Kang, C.M., Lee, I.H., Hyun, C.G., Hong, S.K., and Suh, J.W., Accumulation of S-adenosyl-L-methionine enhances production of actinorhodin but inhibits sporulation in *Streptomyces lividans* TK23, *J Bacteriol*, 185, 592, 2003.
167. Huh, J.H., Kim, D.J., Zhao, X.Q., Li, M., Jo, Y.Y., Yoon, T.M., Shin, S.K., Yong, J.H., Ryu, Y.W., Yang, Y.Y., and Suh, J.W., Widespread activation of antibiotic biosynthesis by S-adenosylmethionine in streptomycetes, *FEMS Microbiol Lett*, 238, 439, 2004.
168. Bertrand, H., Poly, F., Van, V.T., Lombard, N., Nalin, R., Vogel, T.M., and Simonet, P., High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction, *J Microbiol Methods*, 62, 1, 2005.
169. Ginolhac, A., Jarrin, C., Gillet, B., Robe, P., Pujic, P., Tuphile, K., Bertrand, H., Vogel, T.M., Perriere, G., Simonet, P., and Nalin, R., Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones, *Appl Environ Microbiol*, 70, 5522, 2004.

170. Courtois, S., Cappellano, C.M., Ball, M., Francou, F.X., Normand, P., Helynck, G., Martinez, A., Kolvek, S.J., Hopke, J., Osburne, M.S., August, P.R., Nalin, R., Guerineau, M., Jeannin, P., Simonet, P., and Pernodet, J.L., Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products, *Appl Environ Microbiol*, 69, 49, 2003.
171. Khosla, C., McDaniel, R., Ebert-Khosla, S., Torres, R., Sherman, D.H., Bibb, M.J., and Hopwood, D.A., Genetic construction and functional analysis of hybrid polyketide synthases containing heterologous acyl carrier proteins, *J Bacteriol*, 175, 2197, 1993.
172. Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R., and McDaniel, R., Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations, *Biochemistry*, 40, 15464, 2001.
173. McDaniel, R., Kao, C.M., Hwang, S.J., and Khosla, C., Engineered intermodular and intramodular polyketide synthase fusions, *Chem Biol*, 4, 667, 1997.
174. Byrne, B., Carmody, M., Gibson, E., Rawlings, B., and Caffrey, P., Biosynthesis of deoxyamphotericins and deoxyamphoteronolides by engineered strains of *Streptomyces nodosus* , *Chem Biol*, 10, 1215, 2003.
175. McDaniel, R., Ebert-Khosla, S., Fu, H., Hopwood, D.A., and Khosla, C., Engineered biosynthesis of novel polyketides: influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase, *Proc Natl Acad Sci U S A*, 91, 11542, 1994.
176. Miao, V., MF, C.G., Nguyen, K., Brian, P., Penn, J., Whiting, A., Steele, J., Kau, D., Martin, S., Ford, R., Gibson, T., Bouchard, M., Wrigley, S.K., and Baltz, R.H., Genetic engineering in *Streptomyces roseosporus* to produce hybrid lipopeptide antibiotics, *Chem Biol*, 13, 269, 2006.
177. Carreras, C., Frykman, S., Ou, S., Cadapan, L., Zavala, S., Woo, E., Leaf, T., Carney, J., Burlingame, M., Patel, S., Ashley, G., and Licari, P., *Saccharopolyspora erythraea* - catalyzed bioconversion of 6-deoxyerythronolide B analogs for production of novel erythromycins, *J Biotechnol*, 92, 217, 2002.

CHAPTER 8

Metabolic model guided design of *Streptomyces coelicolor* for polyketides production

Irina Borodina, Anna Eliasson Lantz, and Jens Nielsen

Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Authors' contributions

IB – initiated the study, carried out experimental and analytical work, wrote the manuscript
JN, AEL – supervised the work and edited the manuscript

8.1 Abstract

Streptomyces are important industrial organisms, producing more than half of the antibiotics currently in clinical use. We used a model-guided approach to improve actinorhodin production in the model organism *Streptomyces coelicolor* A3(2). A genome-scale model was applied to analyze the changes in metabolic fluxes for single gene-knockouts. The mutants, which were predicted to have higher actinorhodin flux, were constructed and characterized in submerged cultivations. Antibiotics production was not increased significantly in any of the mutants, however, the citrate synthase A deletion mutant and the malate dehydrogenase deletion mutant both secreted pyruvate and acetate, which is an indication of increased pool of a common polyketides precursor – acetyl-CoA. Further work is necessary to direct the excess acetyl-CoA into polyketides biosynthesis.

Keywords

Streptomyces, metabolic model, *in silico* design.

8.2 Introduction

Antibiotics resistance of human pathogens is a rising healthcare problem, particularly in the countries with extensive use of antibiotics (e.g., USA). At the same time the release of new antimicrobial drugs to the market has been decreasing over the past 20 years. Classical discovery methods where soil bacteria are cultivated and tested for inhibition of bacterial growth or other biological activities tend to find mostly compounds that are already known. New post-genomic techniques are emerging; they are based on DNA from non-cultivable species, on activation of silenced biosynthetic clusters and on combinatorial biosynthesis of new drugs by DNA manipulation. These techniques require a good plug&play host to test the activity of the isolated or synthesized biosynthetic clusters. *Streptomyces* species have yielded more than half of the drugs that we currently use and are believed to provide more drug candidates by the usage of post-genomic techniques. Hence, model *Streptomyces* strains are interesting starting points for development of such plug&play systems. *Streptomyces* are soil filamentous bacteria, capable of morphological differentiation and sporulation. Due to their high GC-content (around 70%) and complex regulation of secondary metabolites genes, it is generally the easiest to express *Streptomyces* genes in a *Streptomyces* host.

We chose *S. coelicolor* A3(2) as a model superhost for polyketides production. Our choice was determined by the availability of genomic sequence [1], gene replacement techniques [2], genome-scale metabolic model [3] and extensive physiological characterization of this species [4-6]. *S. coelicolor* is as well a natural producer of polyketide antibiotic actinorhodin and prodigiosins [7, 8]. Genome-scale metabolic models can be used to predict the response of an organism to environmental and genetic perturbations. By performing *in silico* simulations, one can in a matter of minutes test the outcome of all single gene deletions and find out which

deletions that will improve the yield of a desired product or productivity [9, 10]. The method was successfully used to improve lycopene [11, 12] and lactate [13] production in engineered *E. coli*. By using genetic algorithms optimal combinations of several knock-outs/overexpressions can be found [14].
In this study we test whether model-guided strain design can be applied for improving antibiotics-producing *Streptomyces* bacteria.

8.3 Methods

8.3.1 Construction of single deletion mutants

The reference strain was *S. coelicolor* M145 (SCP1$^-$, SCP2$^-$), a prototrophic plasmidless derivative of *S. coelicolor* A3 (2) strain 1147 [1, 15]. The strain was a kind gift from Mervyn Bibb (John Innes Centre, UK). Deletion mutants were constructed using the REDIRECT method [2]. The required *S. coelicolor* cosmids [15], REDIRECT plasmids and *E. coli* strains were kindly supplied by Nick Bird (John Innes Centre, UK). The replacement cassettes carrying apramycin resistance marker, FLP recognition sites and origin of transfer, flanked by 39 and 40 bp upstream and downstream sequences of the to-be-deleted genes were synthesized by PCR. The template for PCR reaction was a fragment of pIJ773 plasmid, cut by *EcoRI* and *HindIII* and purified. The primers were designed to additionally include restriction sites for *NheI*, *XbaI* or *SpeI*, these sites can be used to cut the replacement cassette of the mutated cosmid and to ligate the complementary overhangs, generating "scar" in the place of replaced gene. All the details are given in Table 8-1. For each gene deletion the purified replacement cassette was transformed into BW25113 *E. coli* strain, carrying a plasmid pIJ790 with λ RED recombination genes under control of arabinose-inducible promoter and a cosmids with the gene of interest on it (Table 8-1). After recombination event induced by short cultivation on arabinose, the strains were cured from temperature sensitive plasmid pIJ790 by growth at 37ºC and apramycin-resistant clones were selected. The mutated cosmids were isolated and replacement of the genes with cassette was confirmed by PCR and restriction. The cosmids were transformed into non-methylating *E. coli* ET12567, carrying transfer functions on plasmid pUZ8002. The resulting *E. coli* strains were used to deliver the mutated cosmid into the *S. coelicolor* M145 by conjugation. Apr^{R}Kana^S *S. coelicolor* clones were selected and the gene deletion was confirmed by PCR and restriction; primer sequences used for control PCR are given in Table 8-1. The strains are available on request.

Table 8-1. Primer sequences for construction and confirmation of deletion mutants

Mutant	Cosmid	Primers for deletion[1,2]	Primers for confirmation of deletion
Δ*zwf* (SCO6661)	St5A7	SCO6661_pIJ773_fw: TGATCTTGCTTGCCGTGGTGTCGGTCAGGTCG *SpeI* ATCCTCAACTAGT**TGTAGGCTGGAGCTGCTTC** SCO6661_pIJ773_rv: GACGAAGTCCCTCGACAGCAAGGGAGTTGACG *NheI* GGGAATGGCTAGC**ATTCCGGGGATCCGTCGACC**	SCO6661_ch_fw: TCCTCGTCCGTGACGATGAC SCO6661_ch_rv: TGTCGTACGACGAGGTGGTG
Δ*zwf2* (SCO1937)	StC54	SCO1937_pIJ773_fw: GAGCTCAAGCGCCTCGCTCCCTCGAAGGGCTG *SpeI*	SCO1937_ch_fw: CGGACCTCGACGCGATCGAG SCO1937_ch_rv:

		AGAATTGACTAGT**ATTCCGGGGATCCGTCGACC** SCO1937_pIJ773_rv: TGATCTTGCTGGCCGTGGTGTCCGTGAGGTCG *XbaI* GTCTTCATCTAGA**TGTAGGCTGGAGCTGCTTC**	TCAGCGCGTCGTACGCGTTC
Δ*gap1* (SCO1947)	StC54	SCO1947_pIJ773_fw: CCCTGCTACCGCATCGAGGTGCCTCGCTGAAG *SpeI* TCGATCAACTAGT**TGTAGGCTGGAGCTGCTTC** SCO1947_pIJ773_rv: CACATCGGGCCGGCGTACCAACGAGGAGATCG *XbaI* GTTCGTGTCTAGA**ATTCCGGGGATCCGTCGACC**	SCO1947_ch_fw: CGTCTTCAAGAGGGCTCCTG SCO1947_ch_rv: ATGGGGCGTCTCGATGTCAC
Δ*gap2* (SCO7040)	St2C3	SCO7040_pIJ773_fw: CGCGCTCGCGCCCCGCGTCGCCCCGACCGGAC *SpeI* CGGATCAACTAGT**TGTAGGCTGGAGCTGCTTC** SCO7040_pIJ773_rv: AGCATATGAGCGCCGCCTAGCTCGAAAGATAG *XbaI* AGACGTGTCTAGA**ATTCCGGGGATCCGTCGACC**	SCO7040_ch_fw: GTGGGCACAGCGTCGTCGTC SCO7040_ch_rv: ACGCGCACGATCCCGTACTC
Δ*citA* (SCO2736)	StC57A	SCO2736_pIJ773_fw: GGACTCACCAGACGCGCTGCGGGGGTGCCGTG *SpeI* CGGCTCAACTAGT**TGTAGGCTGGAGCTGCTTC** SCO2736_pIJ773_rv: CCTCGAAGATGAGGCACTACGTCGGTGAGGGA *XbaI* AGACGTGTCTAGA**ATTCCGGGGATCCGTCGACC**	SCO2736_ch_fw: TCTGCTGTCCGGGTCCGTTC SCO2736_ch_rv: GAGGGGGCAGGATGAGGAAG
Δ*mdh* (SCO4827)	St5G8	SCO4827_pIJ773_fw: GGACGCCCCACCGCCAGCTGTCATACGGAGAA *SpeI* CGCCATGACTAGT**ATTCCGGGGATCCGTCGACC** SCO4827_pIJ773_rv: AACCGGAACGGGGCCCGTGCGGGCGTCGGGGC *NheI* GGACTCAGCTAGC**TGTAGGCTGGAGCTGCTTC**	SCO4827_ch_fw: TTGGACGGCTCGGTCCTGTG SCO4827_ch_rv: GCTCCGTTTGCATGCCGAATC
Δ*gdhA* (SCO4683)	StD31	SCO4683_pIJ773_fw: CGCCGAGCGCCTCCCACCGGTCGCTGTATGCG ACCGGTG**ATTCCGGGGATCCGTCGACC** SCO4683_pIJ773_rv: CGCCGGCTCGCGGTGGGGCCCGGTCCGTCGGGT CCGTCA**TGTAGGCTGGAGCTGCTTC**	SCO4683_ch_fw: GGGTAGCGTCGACGGGAG SCO4683_ch_rv: CACGCCGCTACGGTCTTTG

[1] The sequences complementary to replacement cassette ends on plasmid pIJ773 are shown in bold
[2] Restriction sites, which can be used for deletion of the replacement cassette in order to generate clean mutants ("scar" instead of deleted gene), are underlined

8.3.2 Construction of plasmid pIJ6902_act for over-expression of actinorhodin cluster activator

The actinorhodin cluster activator gene *actII-ORF4* was amplified from plasmid pIJ68 [16] in PCR reaction using the following primers: fw_act 5´-TGCAGGCATATGAGATTCAACTTATTG-3´ and 5´-AATTTCTAGAACTACACGAGCACCTTCTC-3´. The fragment was digested with *NdeI* and *XbaI* and ligated into plasmid pIJ6902 [GenBank accession number AJ937361, obtained from Mervyn Bibb, John Innes Centre], which was cut with the same restriction enzymes and dephosphorylated. The ligation products were transformed in DH5α *E. coli* cells. Transformants which grew on apramycin were

screened for the presence of *actII-ORF4* by colony PCR with the same primers as above. The resulting plasmid containing *actII-ORF4* under control of thiostrepton inducible promoter P_{tipA} (pIJ6902_act) was isolated, confirmed by restriction, and transformed into *S. coelicolor* strains by conjugation from *E. coli* ET12567/pUZ8002. The positive transformants of *S. coelicolor* M145 were selected by apramycin and thiostrepton resistance, while the mutants with integrated pIJ6902_act were selected by thiostrepton resistance only because they already possessed apramycin resistance genes from gene deletion procedure.

8.3.3 Cultivation

The *S. coelicolor* spores were densely plated on mannitol soya flour medium [17]. For preparation of spore plates for transformants with plasmid pIJ6902_act, thiostrepton was added to the plates at concentration 25 μg/ml. The spores from each plate were harvested by 1 ml 20% glycerol and used to inoculate 50 ml of 2xYT medium (16 $g \cdot l^{-1}$ tryptone, 10 $g \cdot l^{-1}$ yeast extract, 5 $g \cdot l^{-1}$ NaCl) in 200 ml shake flasks with 20 glass beads of 0.3 mm diameter. The cultures were grown at 30ºC on an orbital shaker with agitation rate 150 rpm for 24 hours. The mycelium was collected by centrifugation at 4,000 g for 5 min, the supernatant was decanted. The mycelium was mixed with the residual liquid and crushed in a crusher, consisting of a glass cylinder and a tight glass plunger. Crashed mycelium was resuspended in 20% peptone to the OD_{340} of 24-25 and stored at -20ºC until further use as inoculum.
500 ml shake flasks with 2 baffles and 20-30 glass beads 4 mm in diameter were used for cultivations. Each shake flask contained 100 ml of defined phosphate limited medium of the following composition: 30 g/L glucose, 4 mM NaH_2PO_4, 100 mM NH_4Cl, 10 mM KCl, 2 mM Na_2SO_4, 2 mM citric acid, 1.25 mM $MgCl_2$, 0.25 mM $CaCl_2$, 5 ml/L trace metal solution, 100 mM 3-(N-morpholino)propanesulfonic acid sodium salt (MOPS), 1 ml/L vitamin solution and 200 μl/L antifoam (Sigma). The trace metal solution contained: 20 mM $FeCl_3$, 10 mM $CuCl_2$, 50 mM $ZnCl_2$, 10 mM $MnCl_2$, 0.02 mM Na_2MoO_4, 20 mM $CoCl_2$, and 10 mM H_3BO_4. The vitamin solution contained (per liter): 0.05 g biotin, 1 g Ca-pantothenate, 1 g nicotinic acid, 25 g myo-inositol, 1 g thiamine-HCl, 1 g pyridoxine-HCl and 0.2 g para-aminobenzoic acid. All solutions were prepared using MilliQ water (Millipore, USA). The salts and antifoam were dissolved in water, pH was adjusted to 7.0 and the medium was autoclaved. Concentrated glucose solution was autoclaved separately and added to the medium after it was cooled down. Vitamins solution was added after autoclavation by sterile filtration. For cultivation of strains carrying integrated plasmid pIJ6902_act 25 μg/ml thiostrepton was added to induce the *actII-ORF4* gene expression. The shake flasks were incubated 30ºC at agitation rate of 150 rpm. The samples were taken aseptically in laminar flow bench.

8.3.4 Biomass estimation by dry cell weight measurements and optical density

To estimate biomass dry weight the whole broth samples of 2.5 - 5 ml volume were filtered through dried and pre-weight 0.45-mm pore size filters (Supor-450, Pall Corporation, Ann Arbor, USA). The biomass on the filter was washed with distilled water and the filter was dried in a microwave oven at 170 W for 20 min, cooled down in a desiccator for 15 min and the mass gain was measured. The optical density was measured at 600 nm using UV-Mini 1240 Shimadzu spectrophotometer (Shimadzu Scientific Instruments, USA). When necessary the samples were diluted with distilled

water to obtain OD_{600} below 0.3, which was within the linear range of the spectrophotometer.

8.3.5 Analysis of pigmented antibiotics

The blue actinorhodin pigment was extracted by vortexing the sample with 3N KOH, the total volume of the mixture being 1.5 ml. The sample volume was from 0.75 ml to 0.075 ml, decreasing for the higher actinorhodin concentrations in the broth. The mixture was centrifuged at 10,000xg for 5 min and the absorbance of the supernatant was measured at 640 nm. The concentration of actinorhodin was calculated using extinction coefficient of 25,320 $M^{-1}cm^{-1}$.

For measuring the red prodigiosin pigments, 0.75 ml of the broth was vortexed with an equal volume of methanol buffered to pH 1.5 with concentrated hydrogen chloride. The extraction was carried out under shaking for 3 hours. The mixture was centrifuged as above and the absorbance of the supernatant was measured at 530 nm. The concentration was found with extinction coefficient of 100,500 $M^{-1}cm^{-1}$.

8.3.6 Analysis of extracellular metabolites by HPLC

Glucose and several extracellular primary metabolites were analyzed by HPLC with an Aminex HPX-87H column (Bio-Rad Laboratories, Hercules, CA, USA) operating at 60°C. The compounds were eluted with 5 mM H_2SO_4 at a flow rate of 0.6 $ml \cdot min^{-1}$. Glucose, ethanol, glycerol, succinate, fumarate, malate and acetate were quantified using a differential refractometer (Waters 410; Millipore, Bedford, MA, USA) whereas pyruvate was quantified using a tunable absorbance detector set at 210 nm (Waters 486; Millipore, Bedford, MA, USA).

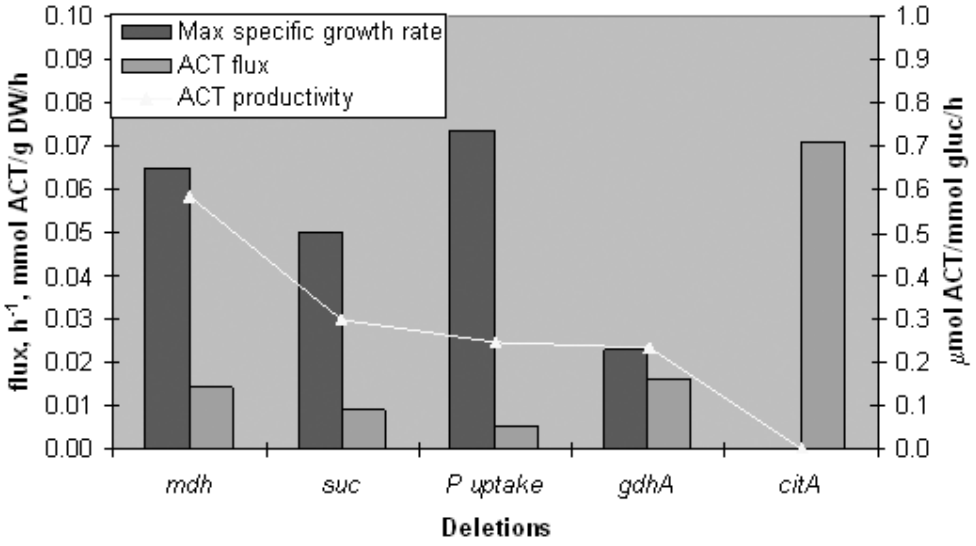
8.4 Results

8.4.1 Identification and construction of optimal single knock-out mutants

To simulate the changes in metabolic fluxes upon single gene deletions we used *S. coelicolor* genome-scale model reconstructed previously [3]. A change was introduced into the model: the transhydrogenase reaction was made irreversible and worked in the direction of NADPH generation from NADH with the usage of membrane protons as described for membrane-bound transhydrogenase from *E. coli* PntAB [18]. *S. coelicolor* genes SCO7623 and SCO7622 encoding correspondingly the α- and β-subunits of transhydrogenase were predicted to contain 5 and 9 transmembrane helices [TMHMM Server, v. 2.0]. Cytoplasmic transhydrogenase, which works in the opposite direction was not identified in the *S. coelicolor* genome using BLAST search with *E. coli* UdhA protein sequence. The outcome of the change was that NADPH could not be converted into NADH by the action of transhydrogenase or some futile transhydrogenations cycles; hence NADPH and NADH were decoupled.

To calculate the distribution of metabolic fluxes in the wild type *S. coelicolor* on minimal mineral medium we introduced constrains for the principal split node ratios as found by C13 metabolic network analysis and performed linear optimization for growth rate considering the experimentally determined glucose uptake rate [6]. The predicted growth rate was in good correspondence to the one determined experimentally (predicted 0.097 h^{-1} versus experimental 0.11 h^{-1}). We then set each flux to zero and calculated the new fluxes using minimization of metabolic adjustment approach (MOMA) [19] (Figure 8-1).

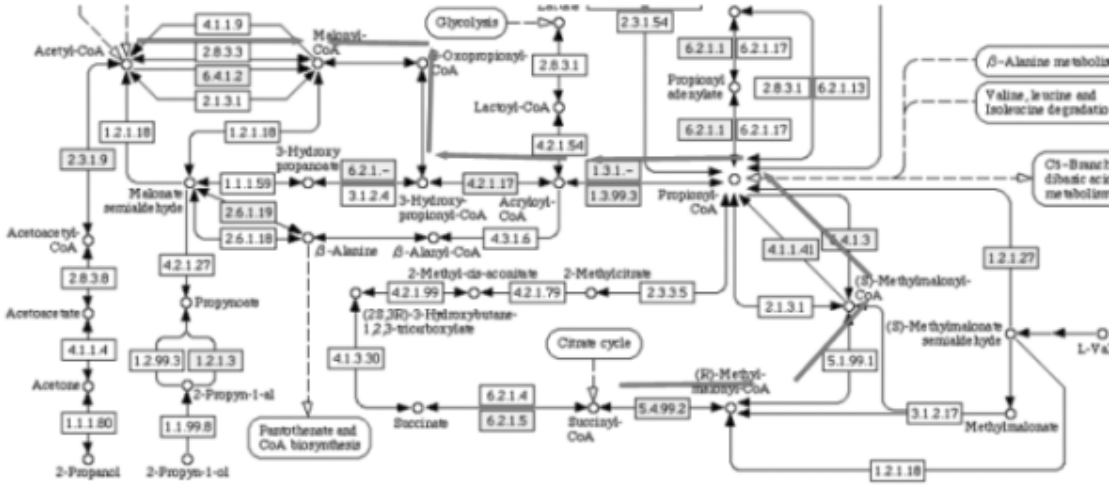
Figure 8-1. Optimal gene deletions (first four by productivity, last by yield) for increase of antibiotics production.



The top four deletions that gave maximal actinorhodin productivity using this approach were: deletion of malate dehydrogenase, deletion of succinate-CoA ligase, decrease of phosphate uptake and deletion of glutamate dehydrogenase. The flux solutions for the mutants explained how actinorhodin increase was obtained. In malate dehydrogenase mutant malate was converted into oxaloacetate through malic enzyme and pyruvate carboxylase, a route, which caused consumption of ATP. There is only one gene coding for malate dehydrogenase (SCO4827), hence the deletion would be easy to make.

Succinate-CoA ligase mutant featured an alternative pathway to malonyl-CoA. Malonyl-CoA is normally synthesized from acetyl-CoA in a single reaction catalyzed by acetyl-CoA carboxylase. In the simulation a 7-reaction route that leads from succinyl-CoA to malonyl-CoA was active (Figure 8-2). It is not highly likely that these reactions can carry a high flux, therefore we chose not to perform this deletion.

Figure 8-2. Alternative pathway for biosynthesis of polyketide precursor malonyl-CoA.

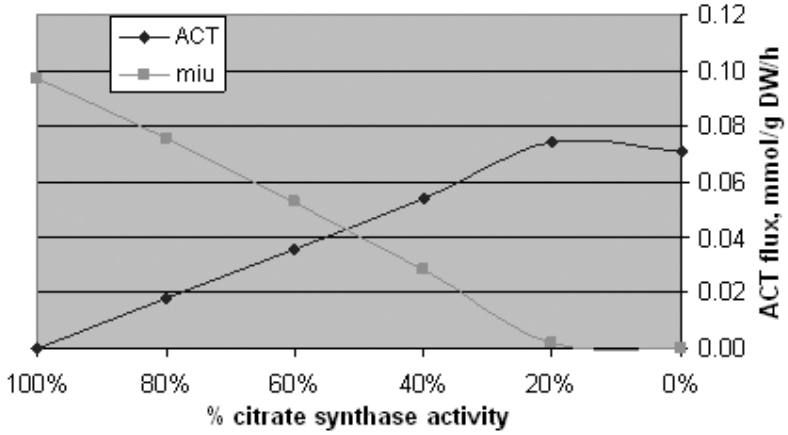


In the third mutant the phosphate uptake by proton symport was deleted and the ATP-dependent phosphate uptake was functioning instead, however with a lower flux. This simulation was equivalent to phosphate limitation: specific growth rate was slower and antibiotics were produced. It is known from experiments that out of several nutrients limitations, carbon-, nitrogen- and phosphate, phosphate-limitation gives the highest actinorhodin titers [5]. Our results show that the positive effect of P-limitation on antibiotics formation is highly influenced by metabolic changes on top of possible transcriptional effects because our model does not account for transcriptional regulation through PhoP boxes and similar.

Another candidate for deletion was glutamate dehydrogenase, which is the main enzyme used for ammonium uptake at high ammonium concentrations in the environment. The enzyme uses NADPH as reducing co-factor per each uptaken ammonium ion and is therefore the most NADPH-consuming enzyme in the cells growing on ammonium (in our simulations it consumes about 60% of all NADPH produced). The model predicted that deletion of the glutamate dehydrogenase would shift the ammonium uptake to the GS-GOGAT system, where ammonium is taken up at the expense of NADH and ATP. This is however not likely to happen in practice because the GS-GOGAT system is inhibited by ammonium and is therefore only active at low ammonium concentrations. According to the genome annotation there is only one glutamate dehydrogenase coding gene – *gdhA* (SCO4683). If this indeed was the only gene, then its deletion would severely impair or even eliminate the growth of our strain on mineral medium with ammonium as nitrogen source. Usually in addition to NADPH-dependent glutamate dehydrogenase, organisms possess NADH-dependent enzyme GdhB, which has lower affinity to glutamate and normally is used for glutamate uptake. In the absence of GdhA, this enzyme was, however, shown to be capable of ammonium uptake [20]. BLAST search with the *Mycobacterium tuberculosis gdhB* gene gave hits to three *Streptomyces* genes: NAD-glutamate dehydrogenase *gdhB* in *S. avermitilis*, putative *gdhB* in *S. clavuligerus* and a hypothetical protein in *S. coelicolor* (SCO2999). The activity of GdhB in *S. clavuligerus* has been confirmed experimentally [21]. SCO2999 contains a motif for bacterial NADH-glutamate dehydrogenase (PFAM05088, Bac_GDH) and has 82% identity to *S. clavuligerus gdhB* and 87% identity to *S. avermitilis gdhB*, it is therefore highly likely that SCO2999 codes for the NAD-dependent glutamate dehydrogenase in *S. coelicolor*. We included the new reaction in the model and simulated what would be the flux distribution in the Δ*gdhA* mutant, if glutamine synthase was inhibited (flux=0) and glutamate dehydrogenase B would be less active than gdhA (half flux of the gdhA in the wild type). The solution was a lower growth rate (0.05 h^{-1} instead of 0.097 h^{-1}) and some actinorhodin production, though less than initially predicted.

The deletion that gave the highest yield of actinorhodin was deletion of citrate synthase, however, the deletion was predicted to be lethal for cells. Obviously, the deletion would eliminate the activity of TCA cycle and direct most of the acetyl-CoA into actinorhodin biosynthetic pathway instead. Citrate synthase is encoded by 3 genes in the genome, of which *citA* (SCO2736) was shown to have the major role in *S. coelicolor* MT1110 [22]. We speculated that deletion of this gene would not eliminate the citrate synthase activity completely, but would reduce the flux through TCA flux and increase the supply of Acetyl-CoA precursor for polyketides (Figure 8-3).

Figure 8-3. Influence of citrate synthase activity on growth and actinorhodin production (computer simulation).



To compare with the model-predicted strategies we chose deletion of *zwf* (SCO6661) and *zwf2* (SCO1937), which were previously shown to improve actinorhodin production in *S. coelicolor* and *S. lividans* [23, 24]. We also chose an empirical strategy – an important glycolytic enzyme, NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (Gap) E.C. 1.2.1.12. Our aim was to block the conversion of glyceraldehyde-3-phosphate (G3P) through 1,3-diphosphateglycerate (13PDG) into 3-phosphoglycerate (3PG) with generation of NADH and ATP and to substitute it with a single-step conversion of G3P into 3PG by NADP-dependent GapN (E.C. 1.2.1.9), generating NADPH and no ATP. We speculated that decreased ATP production and increased supply of NADPH could benefit polyketides production.

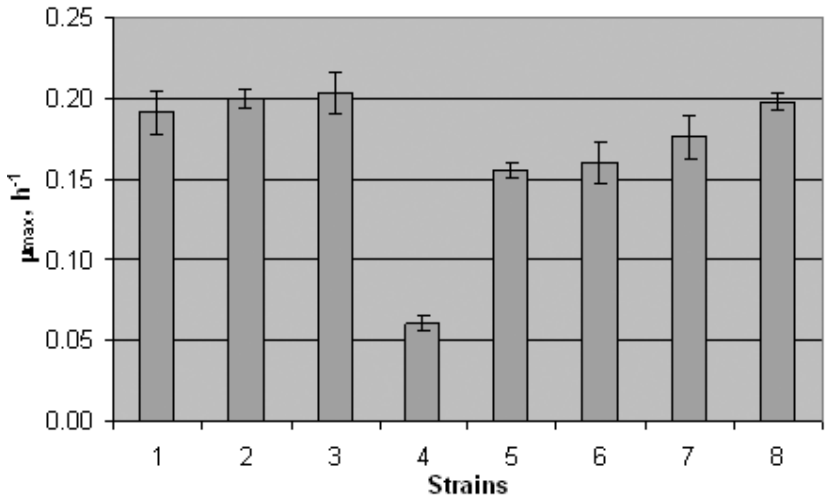
We constructed Δ*mdh,* Δ*gdhA*, Δ*citA*, Δ*gap1*, Δ*gap2*, Δ*zwf*, Δ*zwf2* deletion mutants in *S. coelicolor* M145 by replacing the whole coding sequences with apramycin resistance cassette [2].

8.4.2 Growth

The strains were characterized in shake flasks in defined phosphate-limited mineral medium. Maximal specific growth rate was calculated from the measurements of the optical density at 600 nm in the exponential phase. For all samples taken at 23.5 hours, dry weigh biomass was measured in addition to optical density. There was an 80% correlation between dry weight biomass and OD600 (Appendix 8-1). The μ_{max} calculated for the reference strain M145 was higher than normally measured in fermentor (0.11 h^{-1}), but can be due to evaporation of some water from the culture broth. As evaporation effect would be the same for all the shake flasks, comparisons of the μ_{max} across the strains can be made.

The strains that had significantly ($p<0.05$ in t-test) changed maximal specific growth rates are Δ*gap1*, Δ*gap2* and Δ*gdhA* (Figure 8-4). Glyceraldehyde-3-phosphate dehydrogenase is a key enzyme in glycolysis pathway and should carry a high flux, therefore it is not surprising that *gap* deletion causes growth defects. Deletion of *gap1* decreases the μ_{max} by almost 70%, while knocking out a homologous *gap2* only leads to a 20% decrease. This finding indicates that at least in the exponential growth phase more flux is carried though enzyme Gap1 rather than its isoenzyme Gap2.

Figure 8-4. Maximal specific growth rates of the reference strain M145 and knock-out mutants. Averages of biological triplicates are shown; the whiskers indicate 1 standard deviation. The strains are positioned as following: 1 – M145, 2 – Δ*zwf*, 3 – Δ*zwf2*, 4 – Δ*gap1*, 5 – Δ*gap2*, 6 – Δ*gdhA*, 7 – Δ*citA*, 8 – Δ*mdh*.



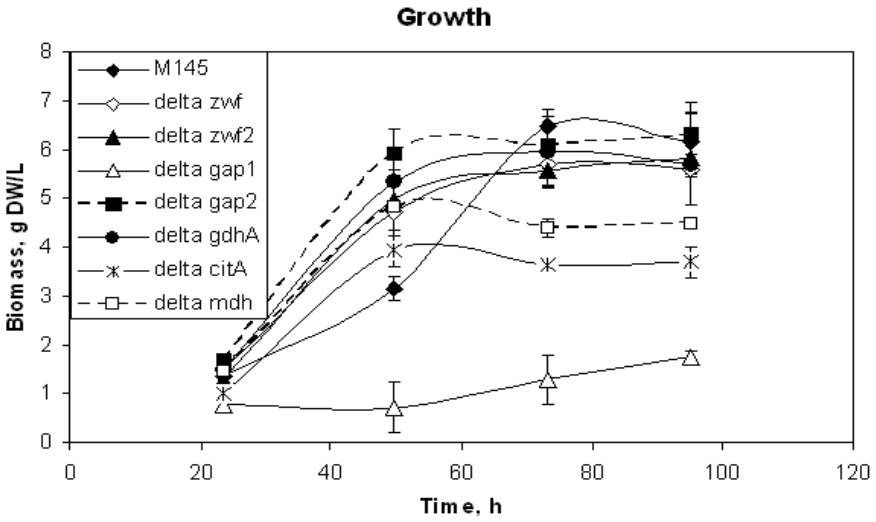
In the strain with glutamate dehydrogenase *gdhA* deleted the μ_{max} is lower by 16%. The fact that the strain is not auxotrophic for glutamate and can grow on ammonium as the sole nitrogen source implies that another ammonium uptake system is functional in the cells. As GC-GOGAT should be repressed at high ammonium concentrations, it is most likely that the ammonium assimilation is carried our by GdhB, though with a slightly lower efficiency.

Surprisingly, the mutants with knock-outs in the important TCA cycle genes *mdh* and *citA* were not severely affected in growth. For citrate synthase, which is expressed from three genes, the isoenzymes might be taking over. For the malate dehydrogenase either the predicted two-step route through malic enzyme and pyruvate carboxylase is quite efficient or some non-annotated additional malate dehydrogenases are present in the genome.

Δ*zwf1* and Δ*zwf2* mutants were not affected in their specific growth rates. Most probably they are mutually interchangeable and can easily take over the pentose phosphate flux, which is not that large in *S. coelicolor* and composes only about 10% of the glucose uptake rate (see Chapter 5).

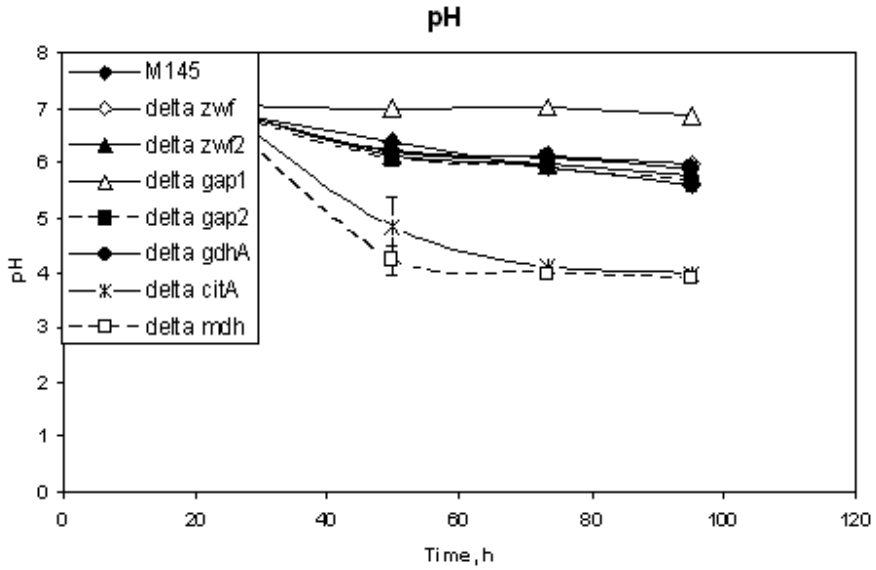
We also monitored the increase of biomass dry weight in the course of cultivation (Figure 8-5). The final biomass concentration was similar for the reference strain and for most of the mutants. Three mutants, however, accumulated much less biomass, these were Δ*gap1*, Δ*citA* and Δ*mdh*. The two later mutants experienced drop of pH to 4-5 already on the second day of cultivation, while for the rest of the strains pH was maintained above 6 until the end of fermentation. It is highly likely that it was the low pH that prevented the citrate synthase and malate dehydrogenase deletion strains from further growth. pH drop must have been due to formation of some organic acids, which could also be expected for mutants in TCA cycle enzymes. We analyzed cultivation broth collected on the 4th day of fermentation from Δ*citA* and Δ*mdh* strains on HPLC and detected pyruvic acid at concentration about 0.7 g/L, acetic acid at 0.9 g/L in Δ*citA* mutant and 0.5 g/L in Δ*mdh* mutant. Additionally we found 1.1-1.2 g/L malic acid and 0.2 g/L fumaric acid.

Figure 8-5. Biomass dry weights of the reference strain M145 and knock-out mutant.



The Δ*gap1* mutant was severely growth deficient; it accumulated less than 2 g DW/L biomass, while biomass level was around 6 g DW/L for non-affected strains. Obviously Gap1 enzyme is important for glycolysis flux not only in the exponential, but also in the later stages of growth.

Figure 8-6. pH change during the shake flask cultivation.

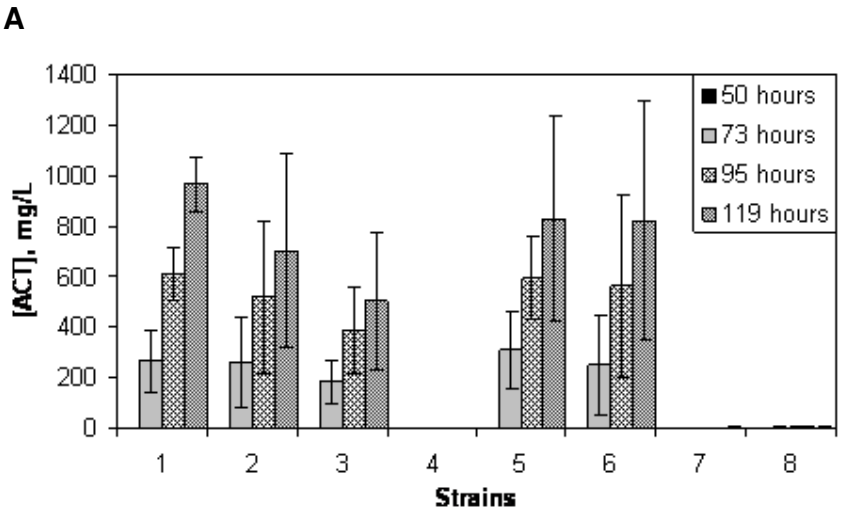


8.4.3 Antibiotics production

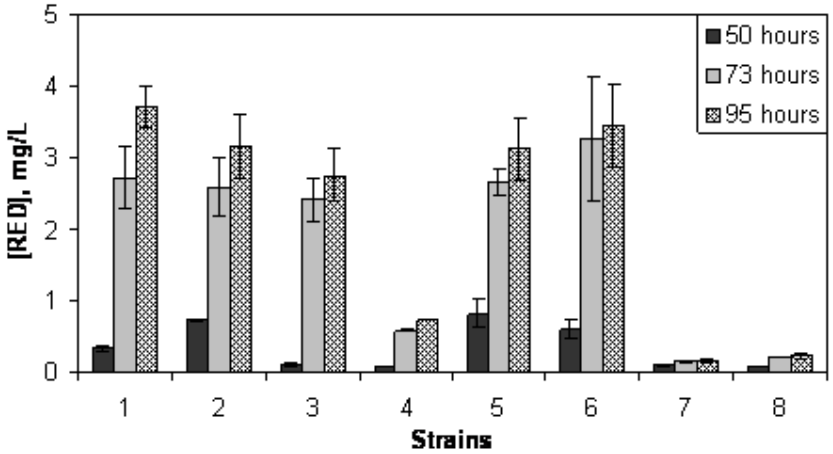
We analyzed concentrations of blue pigmented antibiotic actinorhodin and red pigmented prodigiosin antibiotics. Final antibiotics concentrations were proportional to antibiotics volumetric production rates (data not shown). The variations of actinorhodin production in mutant strains were much higher than in reference strain.

Most commonly one out of three shake flasks would have lower actinorhodin production, while all other parameters as biomass dry weight, pH and prodigiosin production would not differ. It is not clear what caused this high frequency of outliers among the knock-out mutants, considering that mutantions performed by replacement of the whole gene by a resistance cassette should be non-revertable.
The production of both antibiotics did not differ significantly between the reference strain and Δ*zwf*, Δ*gap2* and Δ*gdhA* mutants. If the outlier shake flasks, which produced 3-4 times less actinorhodin than the other shake flasks, were excluded from the analysis, then the average final actinorhodin concentration would be 10-13% higher in the Δ*gap2* and Δ*gdhA* mutants than in the reference strain, though the significance of this change would be low due to the small number of replicates and standard deviation of about 10%.
Antibiotics production was significantly lower in Δ*zwf2*, Δ*gap1*, Δ*citA* and Δ*mdh* strains. For the latter two strains the low antibiotics production can be explained by low pH, which inhibits γ-actinorhodin biosynthesis. At low pH only intracellular actinorhodin is synthesized in small amounts in strain M145 [7]. The major pigment for this strain, however, is the lactone form of actinorhodin - γ-actinorhodin – and it is only produced both intracellularly and in secreted form at neutral pH. To understand the effect of the mutations on antibiotics production, fermentations have to be carried out under pH control. Apparently the capacity of the buffer we used was not sufficient to maintain constant pH and we could not use higher MOPS concentrations because it inhibits cellular growth at concentrations above 100 mM (personal communication with Prashant Bapat).

Figure 8-7. Actinorhodin (A) and prodigiosin (B) concentrations during cultivation of reference strain and single deletion mutants. Averages of biological triplicates are shown, the whiskers indicate 1 standard deviation. The strains are positioned as following: 1 – M145, 2 – Δ*zwf*, 3 – Δ*zwf2*, 4 – Δ*gap1*, 5 – Δ*gap2*, 6 – Δ*gdhA*, 7 – Δ*citA*, 8 – Δ*mdh*.

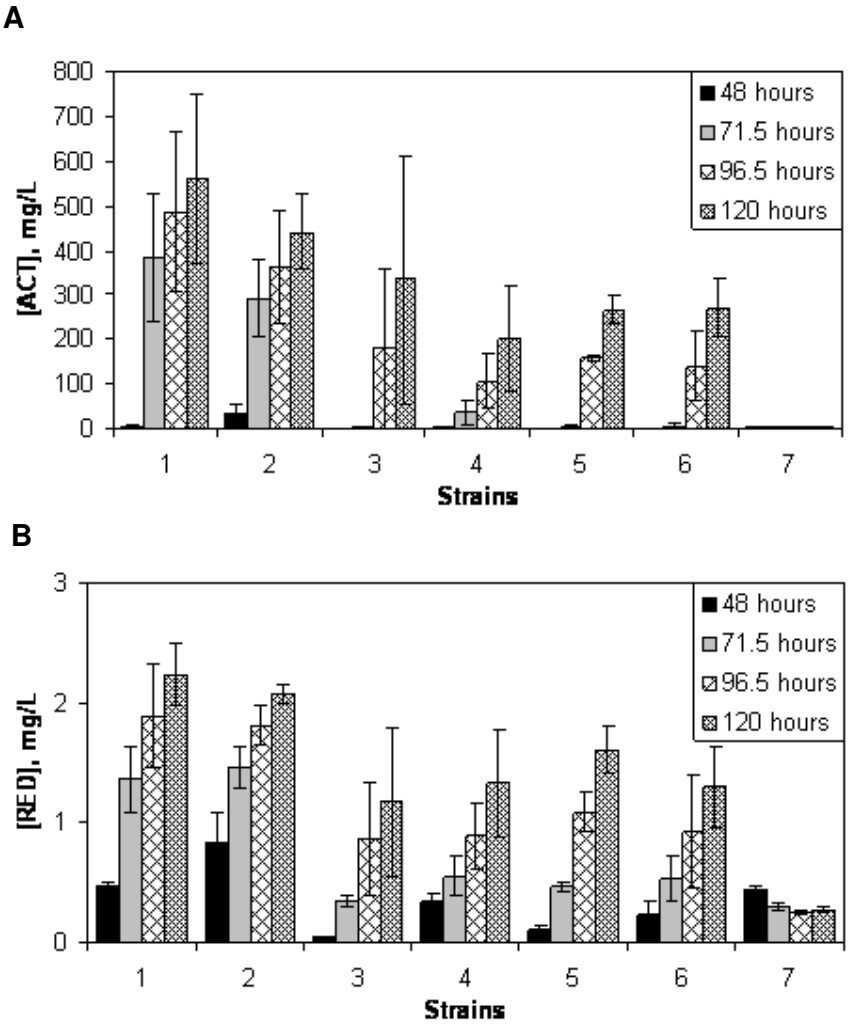


B



By introducing a copy of actinorhodin cluster activator *actII-ORF4* under a strong thiostrepton-inducible promoter P_{tipA} into the M145 strain and its knock-out derivatives, we achieved earlier growth phase-independent actinorhodin production (Figure 8-8). The final titers were, however, not higher than in the strains without the activator overexpressed, and we hypothesize that it was due to a certain inhibition by thiostrepton. We also observed that the lag phase for growth was a bit longer in these than in previous cultivations. The final titers of antibiotics were not improved in any of the deletion strains, but were even significantly lower for many of them (ACT was significantly lower at $p<0.1$ for Δ*gap1*, Δ*gap2*, Δ*gdhA* and Δ*citA* strains and RED for all above listed strains plus Δ*zwf2* strain). The acidification of the fermentation broth during growth of Δ*citA* strain could not be prevented by overexpression of ACT cluster activator and the pH decreased rapidly during the first three days of cultivation as it did before.

Figure 8-8. Actinorhodin (A) and prodigiosin (B) concentrations during cultivation of reference strain and single deletion mutants with integrated plasmid carrying actinorhodin cluster activator gene *actII-ORF4* under control of thiostrepton-inducible promoter. The fermentation medium contained 25 μg/ml thiostrepton. Averages of biological triplicates are shown, the whiskers indicate 1 standard deviation. The strains are positioned as following: 1 – M145, 2 – Δ*zwf*, 3 – Δ*zwf2*, 4 – Δ*gap1*, 5 – Δ*gap2*, 6 – Δ*gdhA*, 7 – Δ*citA*.



8.5 DISCUSSION

Polyketides are an important group of secondary metabolites with a wide pharmaceutical application: polyketides are used as antibiotics, immunosupressants, antiparasitics, cholesterol lowering agents, anti-fungals and anti-cancer drugs [25]. A number of polyketides are made by *Streptomyces* spp. These compounds are synthesized by polyketide synthase enzymes from simple precursors as malonyl-CoA, methylmalonyl-CoA and similar [26]. Malonyl-CoA, which is the precursor for actinorhodin biosynthesis, is formed by carboxylation of acetyl-CoA. It is of interest to alter the cellular metabolism in such a way that the production of polyketides precursors is increased. We used genome-scale metabolic model to predict strategies for increasing supply of acetyl-CoA precursor and NADPH co-factor, which are used in biosynthesis of many polyketides including actinorhodin. We approximated the fluxes in the wild type cells, using the experimentally determined substrate uptake rates, specific growth rate and the split ratios at the major nodes determined from C13 flux analysis [6]. Then we deleted each gene one by one *in silico* and calculated the new flux distribution assuming that the cells will adjust the fluxes minimally to compensate for the mutation (minimization of metabolic adjustment approach [19]). It is also possible to predict the effect of increase of flux

through certain reactions. It is, however, much more difficult to implement such predictions in practice, because overexpression of an enzyme-coding gene does not necessarily lead to increased flux through that enzyme. Even if overexpression of the gene increases the activity of the corresponding enzyme, the flux through it will only increase if the enzyme has a high flux control coefficient and that is rarely the case. We therefore simulated only gene deletions as those most commonly will lead to flux decrease both in simulations and in practice. We did not observe a significant increase in antibiotics final titers for any of the generated mutants, however a number of interesting observations could be made.

Deletion of NADPH-dependent glutamate dehydrogenase was predicted to be one of the strategies for increasing NADPH supply. We found that the Δ*gdhA* strain was capable of growing in defined minimal medium without glutamate supply and had only 16% lower maximal specific growth rate as compared with the reference strain. This is a strong evidence for the presence of an active NAD-dependent glutamate dehydrogenase GdhB, most likely encoded by SCO2999, which has 82% identity to *S. clavuligerus gdhB* and 87% identity to *S. avermitilis gdhB*. Other attempts of changing NADPH and NADH balance through deletion of *ghdA* have been made in yeast *Saccharomyces cerevisiae* [20] and in filamentous fungus *Penicillium chrysogenum*. In both cases it was necessary to concurrently overexpress alternative ammonia uptake routes to avoid growth defects caused by *gdhA* deletion. In yeast the primary metabolism was altered as predicted: glycerol biosynthesis decreased and ethanol production was improved [20]. The outcome with *P. chrysogenum* was however totally unexpected: deletion of *gdhA* fully eliminated production of the antibiotic penicillin (personal communication with Jette Thykær).

Another strategy that we hoped would increase the NADPH production was deletion of glyceraldehyde-3-phosphate dehydrogenase (GAP). We found that Gap1 obviously carried most of the flux because its deletion decreased the μ_{max} with 80% and also lead to a very low biomass accumulation throughout the cultivation. Deletion of *gap2* had a less severe effect on the strain with only 20% decrease in μ_{max} and basically unaltered antibiotics production.

While we did not observe improvement of actinorhodin production in Δ*gdhA*, Δ*gap1* and Δ*gap2* mutants, it still remains to be investigated whether the actually production of co-factor NADPH has been achieved and if yes, then why it did not positively affect antibiotics production.

According to the data of Ryu et al, deletion of glucose-6-phosphate dehydrogenase *zwf1* did not have an effect on actinorhodin production, while *zwf2* deletion increased actinorhoding titers [24]. Perhaps the lack of the positive Δ*zwf2* effect in our study is due to a different *S. coelicolor* strain we are using and due to making fermentations in shake flasks instead of fermentors. However, as we observed positive effect of a mutation under these conditions before for Δ*pfkA2* mutant (Chapter 5), we can conclude that the beneficial influence of *zwf2* deletion is marginal and only true under certain circumstances.

The strains with deletion of citrate synthase A and malate dehydrogenase acidified the fermentation medium to pH 4-5 during the first two days of cultivation, making it impossible to characterize them in shake flasks. Due to time constraints their physiological characterization in fermentors with pH control has not been carried out, but it is a plan for the future. Before the medium acidification occurred we could monitor the exponential growth of these strains and found that their μ_{max} was not significantly changed in comparison to the wild type. In Δ*citA* mutant *citA2* and/or

citA3 might be compensating for the *citA* deletion. The outcome is quite different from *S. coelicolor* MT1110 strain, where deletion of *citA* resulted in glutamate auxotrophy, meaning that CitA activity could not be compensated by other isoenzymes [22]. In Δ*mdh* mutant either the alternative two-step route through malic enzyme and pyruvate carboxylase is fairly efficient or some non-annotated malate dehydrogenase isoenzymes are playing the role. Both Δ*citA* and Δ*mdh* strains produced a range of organic acids as pyruvic, acetic, malic and fumaric acids. Secretion of pyruvic and acetic acids signals that the flux of pyruvate product – acetyl-CoA – into TCA is disturbed. The excess of acetyl-CoA that can not be excreted is converted into acetic acid and the latter is secreted into the extracellular medium. So in both Δ*citA* and Δ*mdh* strains we achieved lower consumption of acetyl-CoA for growth, which most likely lead to higher intracellular acetyl-CoA concentration. A critical question is why this larger pool of acetyl-CoA was not used for actinorhodin biosynthesis, but was rather converted to acetic acid? Whether the fluxes through a node will be redistributed when the node is perturbated depends on the concentrations of the node enzymes and on their affinity to the common substrate. There are many enzymes that compete for acetyl-CoA, like citrate synthase, phosphate acetyltransferase, acetyl-CoA carboxylase, fatty acids synthases, etc. Obviously on deletion of citrate synthase, it was phosphate acetyltransferase that took over the excessive portion of acetyl-CoA. If the affinities of the phosphate acetyltransferase and acetyl-CoA carboxylase to the substrate are in the similar range, then overexpression of acetyl-CoA carboxylase can increase malonyl-CoA biosynthesis and perhaps also polyketides production. The positive effect of acetyl-CoA carboxylase overexpression on actinorhodin production has been shown in the study by Ryu et al. [24]. If, however, the affinity of acetyl-CoA carboxylase is much lower, then deletion of the phosphate acetyltransferase might be necessary to change the distribution of acetyl-CoA flux in favor of polyketides production.

Considering that model-based computer prediction of optimal metabolic strategies is a very young discipline, a few successes have already been made indicating a large potential of the methodology. One of the most prominent was the increase of lycopene production in recombinant *E. coli* overexpressing lycopene biosynthetic genes [11, 12]. The improvement of secondary metabolites production promises though to be more challenging in *Streptomyces*. For the first, secondary metabolites are usually produced in a stationary phase in growth-independent manner, hence, the flux predictions based on growth optimality might not be valid. For the second, the expression of the secondary metabolic genes is often regulated by complex signalling cascades, which can be unintentionally disturbed while metabolically engineering the strains. For the third, pathways for secondary metabolites are often long with many steps and are adjusted to carry low fluxes. The capacity of these pathways has to be improved before engineering the precursor and co-factor supply, which otherwise will not have an effect.

In conclusion, we made the first attempt to use computer simulations, based on stoichiometric metabolic model, to predict metabolic engineering strategies for improved polyketides production in the model *Streptomyces* species, *S. coelicolor*. While we succeeded in increasing the supply of acetyl-CoA in some of the strains, more work is needed to re-direct this polyketides precursor into the desired product.

8.7 Acknowledgements

We are grateful to Mervyn Bibb and Nick Bird for supplying us with the plasmids, cosmids and REDIRECT kit used in this work as well as for advice and guidance in the molecular biology work. We thank Prashant Bapat for preparing stocks of frozen mycelium. The financial support of Technical Universitry of Denmark in the form of PhD grant to IB is gratefully acknowledged.

8.8 References

1. **Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D *et al***. (2002). "*Complete genome sequence of the model actinomycete Streptomyces coelicolor* A3(2)." *Nature*, **417**(6885):141-147.

2. **Gust B, Challis GL, Fowler K, Kieser T, Chater KF**. (2003). "*PCR-targeted Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin." *Proc Natl Acad Sci U S A*, **100**(4):1541-1546.

3. **Borodina I, Krabben P, Nielsen J**. (2005). "*Genome-scale analysis of Streptomyces coelicolor* A3(2) metabolism." *Genome Res*, **15**(6):820-829.

4. **G.P. vW, Krabben P, Traag BA, Keijser BJ, Kerste R, Vijgenboom E, Heijnen JJ, Kraal B**. (2006). "*Unlocking Streptomyces spp. for use as sustainable industrial production platforms by morphological engineering*." *Appl Environ Microbiol*, **72**(8):5283-5288.

5. **Kang SG, Jin W, Bibb M, Lee KJ**. (1998). "*Actinorhodin and undecylprodigiosin production in wild-type and relA mutant strains of Streptomyces coelicolor A3(2) grown in continuous culture*." *FEMS Microbiol Lett*, **168**(2):221-226.

6. **Borodina I, Siebring J, Zhang J, Smith C, Van Keulen G, Dijkhuizen L, Nielsen J**. (2007). "*Antibiotics overproduction in Streptomyces coelicolor* A3(2) mediated by phosphofructokinase A2 deletion." Manuscript in preparation.

7. **Bystrykh LV, Fernandez-Moreno MA, Herrema JK, Malpartida F, Hopwood DA, Dijkhuizen L**. (1996). "*Production of actinorhodin-related "blue pigments" by Streptomyces coelicolor* A3(2)." *J Bacteriol*, **178**(8):2238-2244.

8. **Tsao SW, Rudd BA, He XG, Chang CJ, Floss HG**. (1985). "*Identification of a red pigment from Streptomyces coelicolor A3(2) as a mixture of prodigiosin derivatives*." *J Antibiot*, **38**(1):128-131.

9. **Burgard AP, Pharkya P, Maranas CD**. (2003). "*Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization*." *Biotechnol Bioeng*, **84**(6):647-657.

10. **Pharkya P, Burgard AP, Maranas CD**. (2004). "*OptStrain: a computational framework for redesign of microbial production systems*." *Genome Res*, **14**(11):2367-2376.

11. **Alper H, Miyaoku K, Stephanopoulos G**. (2005). "*Construction of lycopene-overproducing E. coli* strains by combining systematic and combinatorial gene knockout targets." *Nat Biotechnol*, **23**(5):612-616.

12. **Alper H, Jin YS, Moxley JF, Stephanopoulos G**. (2005). "*Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli*." *Metab Eng*, **7**(3):155-164.

13. **Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO**. (2005). "*In silico design and adaptive evolution of Escherichia coli for production of lactic acid*." *Biotechnol Bioeng*, **91**(5):643-648.

14. **Patil KR, Rocha I, Forster J, Nielsen J**. (2005). "*Evolutionary programming as a platform for in silico metabolic engineering*." *BMC Bioinformatics*, **6**:308

15. **Redenbach M, Kieser HM, Denapaite D, Eichner A, Cullum J, Kinashi H, Hopwood DA**. (1996). "*A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome*." *Mol Microbiol*, **21**(1):77-96.

16. **Passantino R, Puglia A, Chater K**. (1991). "*Additional copies of the act*II regulatory gene induce actinorhodin production in pleitropic *bld* mutants of *Streptomyces coelicolor* A3(2)." *J Gen Microbiol*, **137**:2059-2064.

17. **Kieser T, Bibb M, Buttner M, Chater K, D.A. H**. (2000). "*Practical Streptomyces Genetics*." The John Innes Foundation

18. **Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E**. (2004). "*The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli*." *J Biol Chem*, **279**(8):6613-6619.

19. **Segre D, Vitkup D, Church GM**. (2002). "*Analysis of optimality in natural and perturbed metabolic networks*." *Proc Natl Acad Sci U S A*, **99**(23):15112-15117.

20. **Nissen TL, Kielland-Brandt MC, Nielsen J, Villadsen J**. (2000). "*Optimization of ethanol production in Saccharomyces cerevisiae by metabolic engineering of the ammonium assimilation*." *Metab Eng*, **2**(1):69-77.

21. **Minambres B, Olivera ER, Jensen RA, Luengo JM**. (2000). "*A new class of glutamate dehydrogenases (GDH). Biochemical and genetic characterization of the first member, the AMP-requiring NAD-specific GDH of Streptomyces clavuligerus*." *J Biol Chem*, **275**(50):39529-39542.

22. **Viollier PH, Minas W, Dale GE, Folcher M, Thompson CJ**. (2001). "*Role of acid metabolism in Streptomyces coelicolor morphological differentiation and antibiotic biosynthesis*." *J Bacteriol*, **183**(10):3184-3192.

23. **Butler MJ, Bruheim P, Jovetic S, Marinelli F, Postma PW, Bibb MJ**. (2002). "*Engineering of primary carbon metabolism for improved antibiotic production in Streptomyces lividans*." *Appl Environ Microbiol*, **68**(10):4731-4739.

24. **Ryu YG, Butler MJ, Chater KF, Lee KJ**. (2006). "*Engineering of primary carbohydrate metabolism for increased production of actinorhodin in Streptomyces coelicolor*." *Appl Environ Microbiol,* **72**(11):7132-7139.

25. **Weissman KJ**. (2004). "*Polyketide biosynthesis: understanding and exploiting modularity*." *Philos Transact A Math Phys Eng Sci*, **362**(1825):2671-2690.

26. **Hopwood DA**. (1993). "*Genetic engineering of Streptomyces to create hybrid antibiotics*." *Curr Opin Biotechnol*, **4**(5):531-537.

8.9 Web references

http://www.cbs.dtu.dk/services/TMHMM; TMHMM Server, v. 2.0

CHAPTER 9

Metabolic network analysis of *Streptomyces tenebrarius*, a *Streptomyces* with Entner-Doudoroff pathway

Irina Borodina[1], Charlotte Schöller[2], Anna Eliasson[1], Jens Nielsen[1]

[1] Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, building 223, DK-2800 Kgs. Lyngby, Denmark
[2] Scientific Affairs, Alpharma ApS, Dalslandsgade 11, DK-2300 Copenhagen, Denmark

Content of this chapter is published in: Applied and Environmental Microbiology, 2005, 71(5):2294-2302.

9.1 Abstract

Streptomyces tenebrarius is an industrially important microorganism, producing an antibiotic complex, which mainly consists of the aminoglycosides apramycin, tobramycin carbamate and kanamycin B carbamate. When *S. tenebrarius* is used for industrial tobramycin production, kanamycin B carbamate is an unwanted byproduct. The two compounds differ only by one hydroxyl group, which is present in kanamycin carbamate, but is reduced during biosynthesis of tobramycin. ^{13}C metabolic flux analysis was used for elucidating connections between the primary carbon metabolism and the composition of the antibiotic complex. Metabolic flux maps were constructed for the cells grown on minimal medium with glucose or with glucose/glycerol mixture as the carbon source. Addition of glycerol, which is more reduced than glucose, led to a three times reduction of the kanamycin part of the antibiotic complex. The labeling indicated an active Entner-Doudoroff (ED) pathway, which was previously considered to be non-functional in *Streptomyces*. The activity of the pentose phosphate (PP) pathway was low (10-20% of glucose uptake rate). The fluxes through Embden-Meyerhof-Parnas (EMP) and ED pathways were almost evenly distributed during the exponential growth on glucose. During the transition from growth phase to production phase a metabolic shift was observed, characterized by a decreased flux through the ED pathway and increased fluxes through the EMP and PP pathways. Higher specific NADH and NADPH production rates were calculated in the cultivation on glucose/glycerol, which was associated with lower percentage of non-reduced antibiotic kanamycin B carbamate.

Keywords

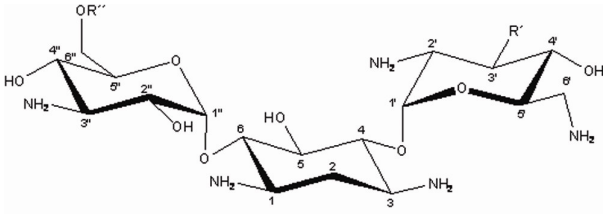
Streptomyces, Entner-Doudoroff pathway, C13 Metabolic Flux Analysis (MFA), tobramycin, kanamycin

9.2 Introduction

The ability of actinomycetes to make secondary metabolites with different useful properties (antibacterial, antitumor agents, immunosuppressants, etc.) is widely exploited in the pharmaceutical industry. Two thirds of the antibiotics produced by microorganisms are made by actinomycetes. Particularly, the *Streptomyces* genus is remarkable in this aspect, representing about 80% of the actinomycete-antibiotics[16].
Antibiotics are formed from specific precursors that are drained from the central carbon metabolism, and over-production of antibiotics therefore requires that the precursors are supplied in sufficient quantities. Improvement of antibiotics production has traditionally been based on random mutagenesis methods, and also in the future these methods will play an important role. However, metabolic engineering enables introduction of rational changes of the central carbon metabolism in order to increase fluxes of precursors and cofactors to antibiotics. Metabolic flux analysis is a valuable tool in guiding the metabolic engineering strategies as it enables rapid phenotypic characterization of different mutants, and through analysis of different mutants one may gain insight into the correlation between antibiotics production and the fluxes through specific branches of the metabolic network.
In this study correlations between the primary and the secondary metabolism in the antibiotics producer *S. tenebrarius*[11,25] was investigated. This species produces the aminoglycoside antibiotic complex nebramycin, consisting primarily of tobramycin

carbamate, kanamycin B carbamate and apramycin[17,18,22]. Tobramycin and kanamycin B carbamates can be hydrolysed into the active forms (tobramycin and kanamycin B, respectively) in a chemical modification step. When *S. tenebrarius* is used for tobramycin production, synthesis of other nebramycin complex components decreases the tobramycin yield and causes problems in the downstream processing. Particularly kanamycin B is difficult to separate from tobramycin due to their chemical similarity. The compounds differ only by one hydroxyl group, which is present in kanamycin B, but is reduced during biosynthesis in tobramycin (Figure 9-1).

Figure 9-1. Chemical structures of tobramycin, tobramycin carbamate, kanamycin B and kanamycin B carbamate.



Strain development and cultivation optimization at the company Alpharma, Denmark, has been used in the past in order to increase the tobramycin carbamate yield and reduce the kanamycin B carbamate content in the complex. Further improvement of the process requires better knowledge of the cellular metabolism, and it is therefore valuable to know whether changes in the antibiotics ratio are connected to the changes in the primary metabolism and how this can be used for directed improvement of the strain. For analysis of the primary metabolism of *S. tenebrarius* metabolic flux analysis was combined with carbon labeling experiments. The cells were grown on ^{13}C-labeled substrate, the labeling of amino acids derived from the cell proteins was analyzed by GC-MS, and through computer modeling the fluxes in the central carbon metabolism were calculated [4,6].

9.3 Materials and methods

9.3.1 Chemicals

All chemicals were of chemical purity degree or higher, purchased from Merck (NJ, USA) or Sigma-Aldrich Co. (MO, USA). The [1-^{13}C]-labeled glucose was from Omicron Biochemicals Inc. (IN, USA).

9.3.2 Strain and culture conditions

Streptomyces tenebrarius strain TD507 (Alpharma ApS, Copenhagen, Denmark) was used as the nebramycin complex producing microorganism. The strain originates from *S. tenebrarius* ATCC 17920 and has been developed by successive mutations either by UV or *N*-methyl-*N'*-nitro-nitrosoguanidine treatments. The strain was stored in 1 ml cryotubes at – 80°C, at a biomass concentration of 9.3 g DW·l^{-1}.

Cultures were grown in batch fermentors with 300 ml growth medium. The medium contained: 2 $g \cdot l^{-1}$ $(NH_4)_2SO_4$, 0.08 $g \cdot l^{-1}$ KH_2PO_4, 0.5 $g \cdot l^{-1}$ $MgSO_4 \cdot 7H_2O$, 10 $mg \cdot l^{-1}$ $ZnSO_4 \cdot 7H_2O$, 10 $mg \cdot l^{-1}$ $FeSO_4 \cdot 7H_2O$, 0.2 $mg \cdot l^{-1}$ $MnSO_4 \cdot H_2O$, 0.02 $mg \cdot l^{-1}$ $CuSO_4 \cdot 5H_2O$, 0.02 $mg \cdot l^{-1}$ $CoCl_2 \cdot 6H_2O$, 2 $g \cdot l^{-1}$ $CaCl_2$, and 1 $ml \cdot l^{-1}$ pluronic as antifoam. As carbon source one of the following was used: 8 $g \cdot l^{-1}$ glucose or 4 $g \cdot l^{-1}$ glucose and 4 $g \cdot l^{-1}$ glycerol.
For labeled substrate cultivations one of the following carbon source combinations was used: 4 $g \cdot l^{-1}$ [1-^{13}C]-glucose and 4 $g \cdot l^{-1}$ naturally labeled glucose or 4 $g \cdot l^{-1}$ [1-^{13}C]-glucose and 4 $g \cdot l^{-1}$ naturally labeled glycerol.
The fermentors were autoclaved with water and pluronic and all other components were added afterwards by sterile filtration. Fermentors were inoculated with 0.3 ml (0.1% v/v) of stock culture. The aeration rate was 1 vvm. For carbon labeling experiments the inlet air was bubbled through a 2 N NaOH solution in order to remove atmospheric carbon dioxide. The pH of the medium was kept at 6.8 by automatic addition of 0.1 N NaOH and the temperature was kept at 37°C throughout the cultivation.

9.3.4 Biomass dry weight

5 ml of culture broth was filtered through a 0.45-µm pore size pre-dried filter (Supor-450, Pall Corporation, Ann Arbor, USA). The filtrate was collected and used for metabolites analysis on HPLC and for analysis of antibiotics. The biomass on the filter was washed twice with distilled water; the filter with biomass was then dried in a microwave oven at 170 W for 20 min, cooled down in a desiccator for 15 min and the mass gain was measured.

9.3.5 Glucose, glycerol and extracellular metabolites analysis

Glucose and several extracellular metabolites were analysed by HPLC with an Aminex HPX-87H column (Bio-Rad Laboratories, Hercules, CA, USA) operating at 60°C. The mobile phase was 5 mM H_2SO_4 at a flow rate of 0.6 $ml \cdot min^{-1}$. Glucose, ethanol, glycerol and succinate were quantified using a differential refractometer (Waters 410; Millipore, Bedford, MA, USA) whereas acetate and pyruvate were quantified using a tunable absorbance detector set at 210 nm (Waters 486; Millipore, Bedford, MA, USA).

9.3.6 Antibiotics analysis

Tobramycin, tobramycin carbamate, kanamycin B, kanamycin B carbamate and apramycin concentrations were analyzed on a Waters HPLC-system (Milford, MA, USA), equipped with a WISP 717 column, and a fluorescence detector (Waters 474), set on 338 nm for excitation and 418 nm for emission measurement. The total eluent flow was 1.2 $ml \cdot min^{-1}$. One liter of the eluent solution contained 175 ml methanol, 3.2 g sodium hexanesulfonate, 12.8 g sodium sulphate, 0.8-1.2 g acetic acid (to obtain pH 3.4), and Milli-Q water (Milli-Q ultrapure water purification system, Millipore, MA, USA). Eluent B differed from eluent A by having twice the amount of methanol. The antibiotics were derivatised by *o*-phthalaldehyde (OPA) and a mercaptan to give fluorescent compounds. The derivatisation was performed at room temperature in a reaction coil installed after the separation column.

9.3.7 GC-MS analysis

From the labeled substrate cultivations samples of 20-30 ml were taken in duplicates for the analysis of cell mass labeling. Samples were filtered through 0.45-µm pore size filter (Supor-450, Pall Corporation, Ann Arbor, USA), the cells were washed twice with distilled water, collected from the filter and stored at –20 ºC until further analysis.

15 mg of wet cell mass was hydrolyzed with 6M HCl at 105 ºC for 24 hours to release free amino acids from protein molecules or for 0.5 hours to free glucose molecules from the cell wall. The hydrolysate was centrifuged at 15,000 x *g* to remove the cell debris, divided into two aliquots and dried at 105ºC. The crude hydrolysates were subjected to derivatisation that rendered them volatile: amino acids were converted to N-ethoxycarbonyl amino acid ethyl esters and N-(N,N-dimethylaminomethylene) amino acid methyl esters and glucose to glucose pentaacetate[4]. Analysis was made by gas chromatography coupled with mass spectrometry as previously described [4]. The signal intensities were corrected for occurrence of natural isotopes in the atoms of the derivative part and in oxygen and nitrogen atoms of the amino acids. The corrected intensities were used for calculating the summed fractional labeling (SFL) of a fragment according to the formula:

$$SFL = \frac{i_0 \cdot 0 + i_1 \cdot 1 + i_2 \cdot 2 + \ldots i_j \cdot m_j \ldots + i_n \cdot n}{i_0 + i_1 + i_2 + \ldots i_j \ldots + i_n} \times 100$$, where i_j is the peak intensity of mass isotopomer m_j and n is the number of carbon atoms in the fragment. The SFL value shows the amount of ^{13}C atoms per 100 fragments.

In order to compare the labeling states of two metabolites with different numbers of carbon atoms carbon normalized labeling (CNL) was calculated as: $CNL = \frac{SFL}{n}$, where SFL is the summed fractional labeling of all the carbons in the fragment and n is the number of carbon atoms in the fragment.

9.3.8 Modelling of metabolic fluxes

A mathematical framework used for quantification of the fluxes was described previously [27,28]. The input to the program included the stoichiometric model, the transitions of carbon atoms in each reaction, the measured SFLs of metabolite fragments, the measured fluxes of substrate uptake and the calculated fluxes towards cell mass. The primary metabolic network of *S. tenebrarius* was reconstructed based on data about carbon metabolism in related species[12]. The fluxes towards cell mass were calculated based on precursor requirements for cell mass biosynthesis in *Corynebacterium glutamicum* [20].

The numeric method (described in more details elswhere[5]) can be summarized as following: first an arbitrary set of fluxes was chosen that fits the stoichiometric constraints and using this set of fluxes the SFLs were calculated. The differences between the measured and the calculated SFLs gave an error, if the error was more than the set value, a new set of fluxes was made and the procedure was repeated until a good fit was obtained. The set of metabolic fluxes that resulted in the best fit were found by an iterative process of error minimization.

9.3.9 Identification of the edd gene

Degenerate primers for the amplification of internal *edd* gene fragment were designed to match regions of high identity in multiple alignments of protein

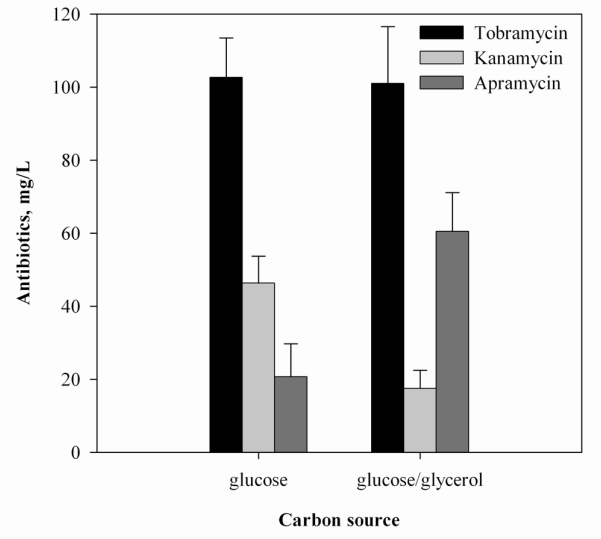
sequences from *Nonomuraea* sp. ATCC 39727 and other bacterial species. The conserved amino acid regions were chosen with a maximum distance from each other in order to amplify the biggest part possible of the gene. To reduce degeneracy of the primers, codon preference of *Streptomyces coelicolor* was considered (Kazusa DNA Research Intitute, Japan [http://www.kazusa.or.jp/codon/]). The primers were extended with 20-nucleotides long tags at the 5'-ends to facilitate subsequent sequencing. PCR reactions were performed using the GC-RICH PCR system (Roche, Mannheim, Germany) and *S. tenebrarius* genomic DNA as template. The coding strand was sequenced at MWG Biotech AG (Reinach, Switzerland) and the sequence was submitted to GenBank (accession number AY618459).

9.4 Results

9.4.1 Influence of carbon source on antibiotics production

The biosynthesis of kanamycin and tobramycin is likely to differentiate only by a reductive reaction, and we therefore investigated how the degree of reduction of the substrate influenced the formation of the antibiotic complex composition in *S. tenebrarius* cultivations. Cultivations were performed with two different carbon sources: pure glucose and an equimolar mixture of glucose and glycerol, where glycerol is a more reduced carbon source than glucose.
Cultivations were performed under controlled conditions on defined minimal medium. The concentrations of antibiotics in the medium were analyzed when the carbon source (-s) was exhausted (Figure 9-2). Tobramycin concentration was not significantly influenced by the nature of the carbon source, while the kanamycin concentration decreased 2.6 times and the apramycin concentration increased 2.9 times when half of the glucose was substituted by glycerol.

Figure 9-2. Final concentrations of antibiotics in *S. tenebrarius* cultivations on glucose and on equimolar mixture of glucose and glycerol. Tobramycin shows the summed concentrations of tobramycin and tobramycin carbamate, kanamycin shows the summed concentrations of kanamycin B and kanamycin B carbamate. Concentrations shown are means of three (glucose) or two (glucose/glycerol) cultivations.



9.4.2 Cultivations on labeled substrate

In order to estimate the fluxes in the central carbon metabolism during growth on glucose and on the glucose/glycerol mixture, carbon labeling experiments were performed. Two batch cultivations were carried out under the same conditions as mentioned above. However, in the glucose cultivation half of the glucose was labeled in the 1st position and the other half was naturally labeled and in the other cultivation a mixture of [1-^{13}C]-glucose and naturally labeled glycerol was used as carbon source. The cells grew at similar specific growth rates and had similar specific glucose uptake rates in the exponential phase of both cultivations (Figure 9-3, Table 9-1).

Figure 9-3. Cultivation profile on minimal medium with [1-^{13}C]-glucose (A) and [1-^{13}C]-glucose/glycerol (B) as the carbon source. The arrows indicate when samples for biomass labeling analysis were taken. Biomass dry weight □, glucose ▲, glycerol ♦, tobramycin and tobramycin carbamate ●, kanamycin and kanamycin carbamate ▽, and apramycin ■.

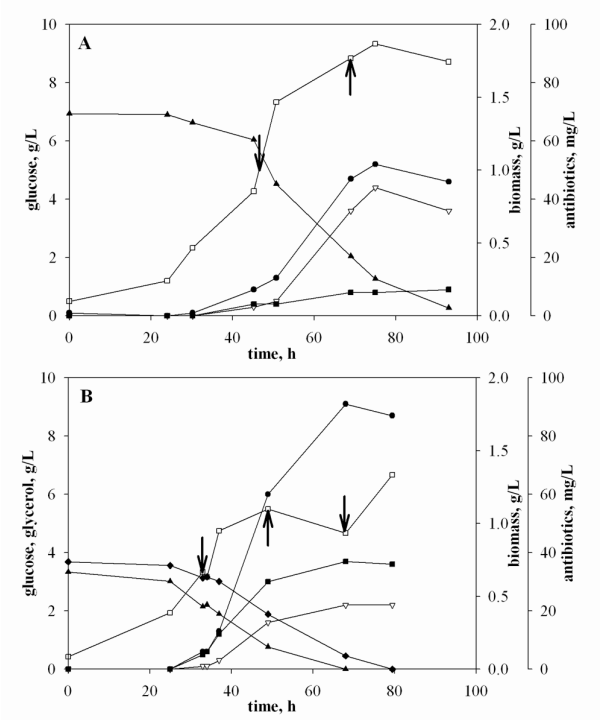


Table 9-1. Some kinetic parameters of growth in the labeled substrate cultivations

Life phase	Parameter	Symbol	Dimension	glucose	glucose/glycerol
Growth phase	Growth rate[a]	μ	h^{-1}	0.061	0.063
	Specific glucose consumption[b]	Y_{xgluc}	g gluc·g^{-1} DW	1.963	1.838
	Specific glucose consumption rate[c]	q_{gluc}	g gluc·g^{-1} DW·h^{-1}	0.120	0.116
	Specific glycerol consumption[b]	Y_{xglyc}	g glyc·g^{-1} DW	-	0.867
	Specific glycerol consumption rate[c]	q_{glyc}	g glyc·g^{-1} DW·h^{-1}	-	0.055
Production phase	Biomass level	x	g DW·l^{-1}	1.750	1.000
	Volumetric glucose consumption rate[d]	v_{gluc}	g gluc·l^{-1}·h^{-1}	0.136	0.094

	Specific glucose consumption rate[e]	q_{gluc}	g gluc·g^{-1} DW·h^{-1}	0.077	0.094
	Volumetric glycerol consumption rate[d]	v_{glyc}	g glyc·l^{-1}·h^{-1}	-	0.082
	Specific glycerol consumption rate[e]	q_{glyc}	g glyc·g^{-1} DW·h^{-1}	-	0.082
	Volumetric antibiotics production rate[d]	v_{antib}	mg·l^{-1}·h^{-1}	3.47	4.08
	Specific antibiotics production rate[e]	q_{antib}	mg·g^{-1} DW·h^{-1}	1.98	4.08

[a] At the exponential growth phase growth rate μ was estimated as the slope of the line, representing natural logarithm of biomass concentration versus time.
[b] The specific glucose consumption Y_{xgluc} was found as the slope of the curve used glucose versus biomass. The specific glycerol consumption Y_{xgly} was found analogically.
[c] Specific glucose consumption rate $q_{gluc}$was calculated as the result of μ and Y_{xgluc} (analogically, $q_{glyc}=\mu \cdot Y_{xgly}$).
[d] For production phase standard approach (c) was not suitable, because there was no net biomass production. From the curve of glucose concentration versus time volumetric glucose consumption rate was calculated instead. Volumetric glycerol consumption rate and volumetric antibiotics production rate were found analogically.
[e] Specific glucose consumption rate was found by dividing volumetric rate of glucose consumption by the approximate level of biomass during production phase. The same method was used for calculation of glycerol specific consumption rate and antibiotics specific production rate.

Even though the final tobramycin concentration was lower in the fermentation with glucose as carbon source when a labeled carbon source was used (Figure 9-3) compared with a non-labeled carbon source (Figure 9-2), the change in the relative production of the different components of the antibiotic complex was the same upon shift of carbon source in the two sets of experiments. During the production phase, which starts when the cells are still growing exponentially but continues into the growth decelleration phase, the substrate uptake continued to be higher in the cultivation with glucose/glycerol as carbon source and the specific rate of antibiotics production was two times higher. The tendency of higher tobramycin and lower kanamycin production in cultivation on glucose/glycerol was preserved in labeled substrate fermentations. During the fermentation, biomass samples were taken at the end of the exponential growth phase (47 h for glucose fermentation and 33 h for glucose/glycerol fermentation) and during the production phase (70 h for glucose fermentation and 49 and 68 h for glucose/glycerol fermentation) and the summed fractional labelings (SFLs) of derivatized proteinogenic amino acids were determined by GC-MS analysis (Table 9-2).

Table 9-2. Measured summed fractional labelings*

Monomer		Derivat. method	m/z	C atoms in measured fragment	Precursor (-s)	C atoms in precursor (-s)	Summed fractional labellings SFLs, %				
							glucose		glucose/glycerol		
							47 h	70 h	33 h	49 h	68 h
Glucose (broth)	Glc	pentaacet	331	1, 2, 3, 4, 5, 6	-	-	61.7±0.1	-	-	-	-
Glucose (biomass)	Glc	pentaacet	331	1, 2, 3, 4, 5, 6	G6P	1, 2, 3, 4, 5, 6	60.1±3.1	63.2±1.1	85.1±2.8	82.1±3.1	76.6±0.5
Alanine	Ala	DMFDMA	99	2, 3	PYR	2, 3	10.9±0.0	15.7±0.0	9.3±0.2	11.6±0.0	10.1±0.1
Alanine	Ala	ECF	116	2, 3		2, 3	10.8±0.3	16.6±1.0	9.6±0.2	13.0±0.4	9.9±0.1
Valine	Val	DMFDMA	127	2, 3, 4, 5		2, 2, 3, 3	20.4±0.2	29.5±0.0	17.0±0.0	21.4±0.0	20.2±0.1
Valine	Val	ECF	144	2, 3, 4, 5		2, 2, 3, 3	20.7±0.2	30.4±0.8	17.6±0.7	22.9±0.9	19.8±0.2
Alanine	Ala	DMFDMA	158	1, 2, 3		1, 2, 3	25.6±0.2	26.3±0.7	31.8±0.3	28.3±0.2	25.2±0.2
Valine	Val	DMFDMA	186	1, 2, 3, 4, 5		1, 2, 2, 3, 3	37.2±0.0	42.0±0.0	46.0±0.1	43.7±0.2	43.2±1.3

Valine	Val	DMFDMA	143	1, 2		1, 2	16.9±0.1	14.0±0.4	25.1±0.2	20.4±0.2	19.2±0.0
Leucine	Leu	ECF	158	2, 3, 4, 5, 6	PYR + ACCOA	2, 2, 3, 3 + 2	29.1±0.6	39.6±1.5	23.9±1.3	30.2±1.4	25.6±0.2
Isoleucine	Ile	ECF	158	2, 3, 4, 5, 6	OAA + PYR	2, 3, 4 + 2, 3	26.6±0.4	36.8±1.8	21.3±1.1	24.2±0.6	23.9±1.9
Lysine	Lys	ECF	156	2, 3, 4, 5, 6		2, 3, 4 + 2, 3	31.5±0.5	41.1±0.2	28.2±0.1	33.0±0.3	29.0±0.2
Aspartic acid	Asp	ECF	188	2, 3, 4		2, 3, 4	20.3±1.2	29.8±0.4	20.7±0.3	23.5±0.6	17.5±0.2
Threonine	Thr	ECF	146	2, 3, 4		2, 3, 4	19.3±1.2	23.9±0.0	19.8±0.5	21.7±0.2	19.6±1.7
Aspartic acid	Asp	DMFDMA	216	1, 2, 3, 4	OAA	1, 2, 3, 4	31.3±0.4	36.6±0.8	49.3±6.2	42.7±1.3	53.0±5.3
Aspartic acid	Asp	DMFDMA	115	2		2	6.0±0.3	8.4±0.2	8.7±2.5	8.1±0.0	8.7±1.0
Threonine	Thr	ECF	175	1, 2		1, 2	13.7±0.7	16.2±0.2	17.0±0.5	15.7±0.3	14.7±0.8
Glycine	Gly	DMFDMA	144	1, 2		1, 2	4.4±0.1	5.0±0.2	6.3±0.1	5.5±0.0	6.5±0.3
Glycine	Gly	ECF	175	1, 2		1, 2	5.8±0.4	7.0±1.1	7.9±0.2	7.7±0.2	6.4±0.5
Serine	Ser	ECF	175	1, 2	3PG	1, 2	4.0±1.7	3.6±3.2	3.1±0.8	3.4±1.3	4.3±1.2
Glycine	Gly	DMFDMA	85	2		2	2.4±0.1	2.2±0.0	4.1±0.0	2.9±0.2	3.7±0.8
Serine	Ser	ECF	132	2, 3		2, 3	14.4±0.2	17.7±0.3	10.0±0.2	12.0±0.7	10.5±0.5
Proline	Pro	ECF	142	2, 3, 4, 5		2, 3, 4, 5	23.2±0.3	30.3±0.8	17.5±0.4	23.7±2.1	22.2±0.4
Glutamic acid	Glu	DMFDMA	230	1, 2, 3, 4, 5	AKG	1, 2, 3, 4, 5	34.8±0.7	46.9±0.4	43.0±4.7	40.1±0.1	36.1±2.3
Glutamic acid	Glu	DMFDMA	143	1, 2		1, 2	17.6±0.9	22.0±1.4	24.9±2.8	22.1±0.8	18.5±3.1
Phenylalanine	Phe	DMFDMA	143	1, 2	PEP	1, 2	3.5±0.3	4.1±0.1	4.0±0.1	3.9±0.2	3.5±0.2
Phenylalanine	Phe	ECF	192	2, 3, 4, 5, 6, 7, 8, 9	E4P + PEP	1, 2, 3, 4 + 2, 2, 3, 3	40.0±1.8	49.6±0.1	31.1±0.5	38.6±0.6	36.5±0.0

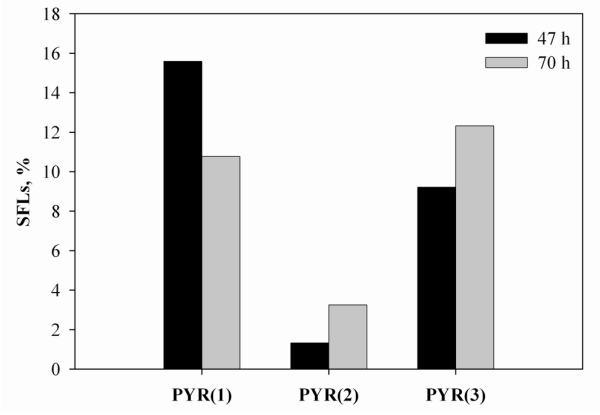
* Mean values and standard deviations between samples

9.4.3 Labeling analysis

9.4.3.1 Pyruvate labeling

Pyruvate is a central carbon metabolite, which labeling pattern can be deduced from the labeling patterns of the amino acids alanine and valine as these have pyruvate as precursor. The amount and diversity of the fragments measured allowed calculation of the fractional labeling of all three carbon atoms in pyruvate (Figure 9-4, Table 9-2).

Figure 9-4. Labeling of pyruvate atoms in the cultivation on glucose. The SFL of the pyruvate 2nd and 3rd carbon atoms was calculated as an average between SFLs of Ala116, Ala99, half SFLs of Val127 and half SFLs of Val144. The labeling of the 1st pyruvate carbon atom was found as average between differences: Ala158 – PYR(2, 3) and Val186 – PYR(2, 3). The SFL of pyruvate 2nd carbon was calculated as difference between Val143 and PYR(1). Eventually the SFL of the third position was found as difference between SFL of PYR(2, 3) and PYR(2).

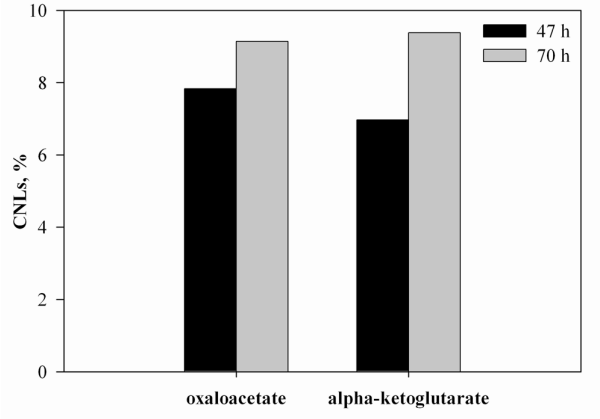


When [1-^{13}C]-glucose is metabolized into pyruvate via the Embden-Meyerhof-Parnas (EMP) pathway, 50% of the pyruvate formed is labeled in position 3. If glucose is metabolized via the Entner-Doudoroff (ED) pathway 50% of the pyruvate formed will be labeled in position 1. This pathway is less energetically favorable than the more common EMP pathway. When the pentose phosphate (PP) pathway is active, the labeled carbon atom is converted to carbon dioxide and the resulting pyruvate does not carry any ^{13}C atoms.
For the cultivation on glucose the labeling of the first pyruvate position was much higher than the natural labeling (1.1%), clearly pointing to an active ED pathway. The EMP pathway was, however, also active as could be judged from the labeling of the third carbon atom in pyruvate. When comparing pyruvate labeling at the exponential growth phase and after onset of the production phase an interesting change in the incorporation pattern was observed: the labeling of the 1st atom decreased, while the labeling of the 2nd and 3rd atoms increased. This can be interpreted as a metabolic shift from ED towards EMP activity when the cells passed from growth to secondary metabolites production. It is possible to judge about activity of the oxidative branch of the PP pathway in the glucose cultivation from the data of glucose labeling in the medium. The SFL of glucose atoms in the medium was equal to 61.7±0.1%. If glucose was metabolized solely through the EMP and ED pathways, the summed fractional labeling of pyruvate should be about 31% (one mole of glucose gives 2 moles of pyruvate). However, at 47 hours the measured SFL was 26.1%, and this SFL value implies that glucose is partly metabolized through the oxidative PP pathway.

9.4.3.2 Tricarboxylic acid (TCA) cycle metabolites

The average labeling of carbon atoms (CNL) in the two TCA cycle metabolites oxaloacetate and α-ketoglutarate was found from the fragments of aspartate (Asp216) and glutamate (Glu230), respectively (Figure 9-5). The CNL values were equal as expected because there is a constant interconversion between oxaloacetate and α-ketoglutarate atoms in the TCA cycle. The increase of labeling with time can be explained by the increase of the labeling of the 2nd and 3rd carbon atoms of pyruvate, which enters the TCA cycle in the form of acetyl-CoA.

Figure 9-5. Labeling of the TCA cycle metabolites: oxaloacetate and α-ketoglutarate in cultivation on glucose. The labelings are normalised in respect to carbon atoms.



9.4.4 Computer simulations of fluxes at the exponential phase

Based on the knowledge about the primary metabolism of streptomycetes [12], a metabolic model was constructed for *S. tenebrarius*. Using this model the fluxes through the different pathways were estimated (see material and methods) for cells in the exponential growth phase, where the system can be considered to be in pseudo-steady state (Figure 9-6).

The irreversibility of fructose-6-phosphate conversion into fructose-1,6-biphosphate (which was initially imposed in the model) resulted in a poor fit of the glucose-6-phosphate labeling in the fermentation on the two carbon sources (104% predicted against 85.1% measured). One factor that could explain this deviation would be that the model predicts too low flux through the oxidative PP pathway activity, but this option was excluded due to an almost complete accumulation of the label from glucose in pyruvate. Thus, there must be a route resulting in dilution of the labeling at the level of glucose-6-phosphate and the only possible option is reversibility of the reactions converting fructose-6-phosphate into glyceraldehyde-3-phosphate and dihydroxyacetone-phosphate. Dihydroxyacetone-phosphate is an intermediate product of glycerol catabolism, and part of it will therefore be only naturally labelled. Hence activity of fructose-1,6-biphosphatase was included in the model for the simulations of growth on glucose/glycerol.

9.4.4.1 Glycolysis pathways

The distribution of fluxes around the G6P branch point was quite different for the two different carbon sources.

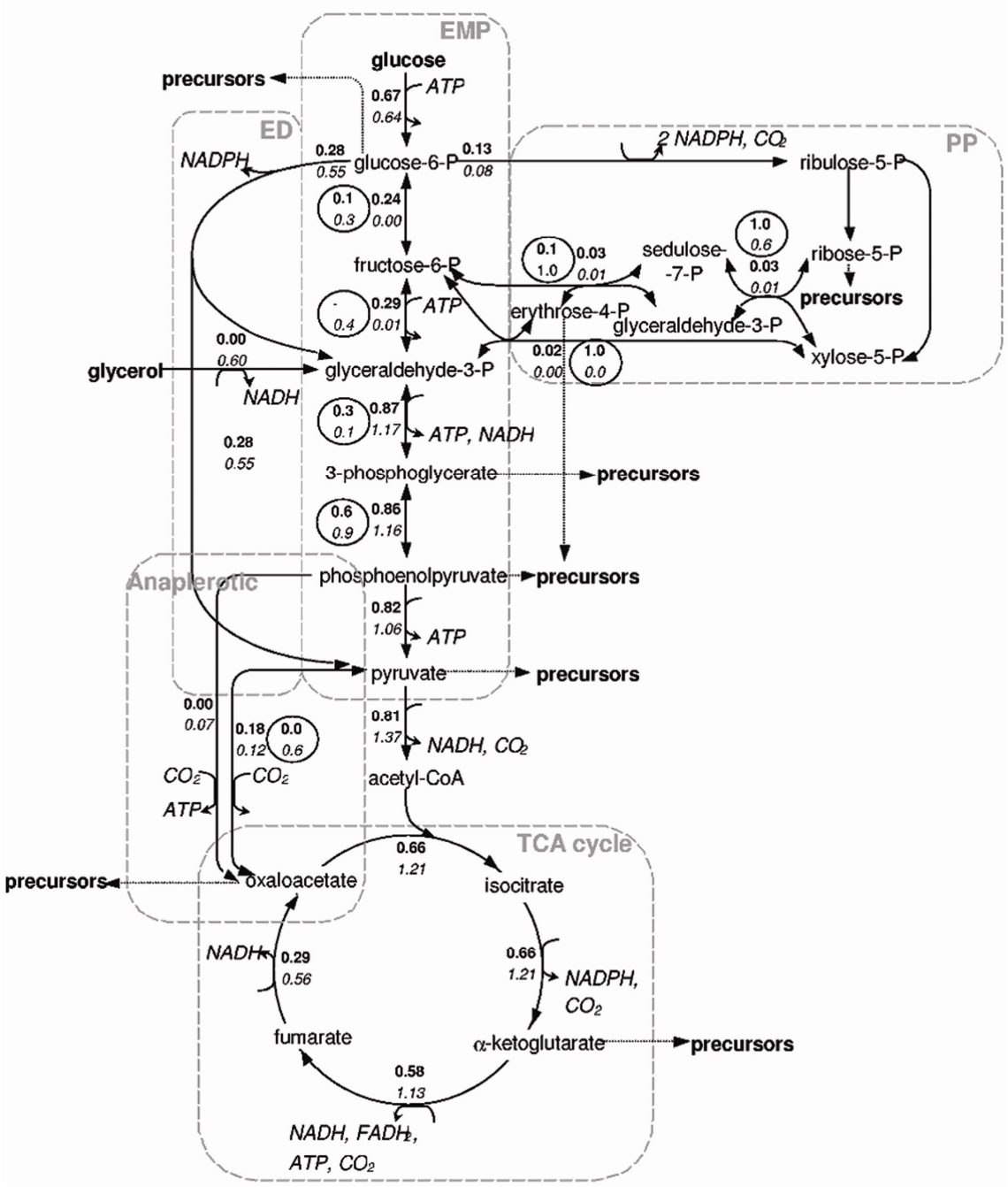
In the cultivation on glucose the ED and EMP pathways were equally important for glucose conversion, while about 20% of the glucose was shunted through the PP pathway (some G6P was channeled into biomass). Less than half of the carbon that was oxidized in the first step of the PP pathway returned to the EMP pathway, whereas the remaining glucose that entered the PP pathway was used as building blocks for nucleotides and aromatic amino acids.

In the glucose/glycerol cultivation most of the G6P (85.1%) was directed towards the ED pathway. The remaining 12.4% went to the PP pathway, which was just sufficient to satisfy the needs for precursor metabolites required for biomass synthesis. Thus, in this cultivation the main role of the PP pathway was supply of precursors and not the formation of NADPH. Despite the presence of labeling in the 3rd position of pyruvate there was no net flux of G6P into the EMP pathway. This can be explained by high exchange rates in the reactions leading from G6P to G3P. The lower PP pathway activity in the glucose/glycerol cultivation is also reasonable from a biological point of view: the ED pathway supplied the cell with NADPH required for growth, and it is therefore not necessary to use the PP pathway for supply of NADPH and this pathway is exclusively used for supply of biomass precursors.

9.4.4.2 Anaplerotic pathways

Out of the two anaplerotic pathways that were included in the model, pyruvate carboxylation was chosen as the one that gave the best fit to the labeling patterns in the glucose cultivation. This was surprising because PEP carboxylase is regarded as the enzyme primarily responsible for TCA cycle metabolite replenishing in *Streptomyces* [2,8].

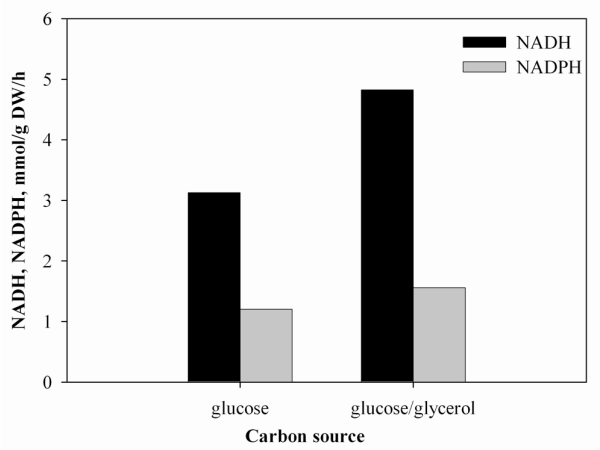
Figure 9-6. The metabolic fluxes as calculated in computer simulation. The values are averages between fluxes obtained in two simulations. Numbers in bold are fluxes for the glucose cultivation, numbers in italic are fluxes for the glucose/glycerol cultivation. All fluxes are given in mmol·g^{-1} DW·h^{-1}. The numbers in circles show exchange coefficients for reversible reactions, calculated as: $\frac{MIN(v\_forward, v\_reverse)}{100 + MIN(v\_forward, v\_reverse)}$, where $v\_forward$ is forward flux, $v\_reverse$ is reverse flux. The function $MIN(v\_forward, v\_reverse)$ gives the smaller flux (forward or reverse) as the result.



9.4.5 NADH, NADPH production

With respect to the availability of cofactors for synthesis of antibiotics it is important to know how much NADH and NADPH is synthesized and consumed by the cell. In this case it was interesting to know, whether more reducing equivalents were produced per g dry weight per hour when glucose was partly substituted by glycerol. The cells grew at approximately the same specific growth rates during both cultivations and the cell compositions were assumed similar. Consequently, the specific fluxes of NADH production and NADPH consumption during biomass synthesis were considered similar for both cultivations. Therefore, the main difference was determined by the amount of cofactors produced, which could be calculated from the estimated fluxes and the stoichiometric coefficients for NADH, NADPH in each of the reactions included in the model (Figure 9-7).

Figure 9-7. Specific reducing cofactors production in the cultivation on glucose and on glucose/glycerol.



The specific net production rate of NADH in the exponential growth phase was higher in the cultivation on glucose/glycerol (4.8 mmol·g^{-1} DW·h^{-1}) than in the cultivation on glucose (3.1 mmol·g^{-1} DW·h^{-1}). This can be explained by a higher carbon uptake rate, 5.66 C-mmol substrate·g^{-1} DW·h^{-1} in the glucose/glycerol cultivation compared to 4.00 C-mmol substrate·g^{-1} DW·h^{-1} in the cultivation on pure glucose. Another reason is the higher degree of reduction of glycerol, which results in the formation of more NADH during catabolism.

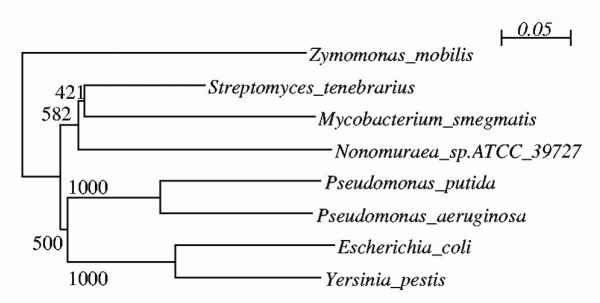
The specific net production rate of NADPH was slightly higher in the cultivation on glucose/glycerol, 1.6 mmol·g^{-1} DW·h^{-1}, compared with the cultivation on glucose, 1.2 mmol·g^{-1} DW·h^{-1}. Dividing these values by the specific growth rate (app. 0.06 h^{-1}), one obtains the specific NADPH formation for the two cultivations, 27 and 20 mmol·g^{-1} DW, respectively. The requirement for NADPH for biomass biosynthesis varies for different organisms; for *P. crysogenum* it was calculated to be 8.5 mmol·g^{-1} DW[21], for *C. glutamicum* the value was 14.9 mmol·g^{-1} DW[20]. From this it is clear that the predicted NADPH formation rate in *S. tenebrarius* can satisfy the biomass biosynthesis.

Judging from the substrate consumption rate (which was twice as high in the glucose/glycerol cultivation), the tendency of higher specific production rates of the reduced cofactors in the glucose/glycerol cultivation should also hold for the stationary phase, when the antibiotics were produced.

9.4.6 Identification of the edd gene

In order to further confirm the presence of an ED pathway in *Streptomyces tenebrarius*, a 1.272 kb-fragment of the *edd* gene, encoding the ED pathway specific enzyme 6-phosphogluconate dehydratase, was amplified. The amplification was performed with degenerate primers on genomic DNA of *S. tenebrarius* as the template. 675 base pairs were successfully sequenced. The putative amino acid sequence showed 70% identity over 225 residues to Edd from *Pseudomonas putida*, *P. aeruginosa*, and *E. coli*. In a clustering analysis the sequence grouped with Edd from other actinomycetes with proven ED pathway activity – *Nonomuraea* sp. ATCC 39727 and *Mycobacterium smegmatis* (Figure 9-8).

Figure 9-8. Unrooted phylogenetic tree derived by neighbor-joining analysis of Edd amino acids sequences from different bacteria species with proven Entner-Doudoroff pathway presence. The bootstrap values on the branches indicate the number of times a given branch appeared in 1000 bootstrap replications. The sequences were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov). ClustalX 1.8 was used for making multiple sequence alignment and for construction of the tree. The tree was drawn with Njplot.



9.5 Discussion

Actinomycetes are an important group of bacteria for the biotechnological industry, mainly due to their ability to produce a wide range of bioactive metabolites. The secondary metabolism of actinomycetes has been studied intensively, but the primary metabolism of most actinomycetes is, however, poorly characterised. In this study, a correlation between the primary metabolism and the aminoglycoside production in *S. tenebrarius* was sought.

9.5.1 Detection of Entner-Doudoroff pathway

The Entner-Doudoroff pathway for carbohydrate metabolism[9] is most common in Gram-negative bacteria, though it has also been encountered in Gram-positive bacteria and even in some eucaryotes[7]. The ED pathway has never been described in *Streptomyces* before and is unusual for actinomycetes altogether. To the best of our knowledge the only actinomycetes found to possess an ED pathway are *Mycobacterium smegmatis*[1] and *Nonomuraea* spp. ATCC 39727[10]. The *edd* gene has not been found in the sequenced genomes of the two *Streptomyces* species (*S. coelicolor* and *S. avermitilis*)[10]. The ED pathway activity was shown in *S. tenebrarius* by labeling experiments, and the presence of the characteristic gene *edd* was confirmed by PCR and subsequent sequencing. The activity of the pathway could not be verified using the enzymatic assay of Budgen and Danson[3], because of a high non-specific NADH-oxidative activity of the cell extract (data not shown).

Labeling experiments indicated simultaneous activity of glycolysis and gluconeogenesis, which basically means that there is an inefficient energy metabolism. Possibly the ED pathway functions in a cyclic mode as in pseudomonads, where part of the dihydroxy-acetone phosphate produced in the ED pathway is recycled back to fructose-6-phosphate and then to gluconate instead of being directly processed into pyruvate[7]. An alternative explanation is that gluconeogenesis is active only during metabolism of glycerol. Analogy could again be drawn with pseudomonads, where disruption of the *edd* gene is deleterious for growth on glycerol[19], which shows that glycerol is first turned into glucose-6-phosphate and then metabolized via the ED pathway. Lower *S. tenebrarius* biomass yield during growth on glucose/glycerol (0.34 g DW/g substrate versus 0.50 g DW/g glucose) supports the latter hypothesis. Normally the formation of a futile cycle due to fructose-6-phosphate and fructose-1,6-diphosphate interconversion is avoided by microorganisms. For instance *E. coli* operates the ED pathway only in a linear mode and pseudomonads are devoid of 6-phosphofructokinase and therefore cannot metabolise carbohydrates via the EMP pathway. It has been shown that when 6-phosphofructokinase was introduced into *Alcaligenes eutrophus* (related to pseudomonads), the growth of the organism was impaired. However, when 6-phosphofructokinase was introduced into an Edd^- mutant, the organism successfully used the EMP pathway and grew normally[26].

9.5.2 Metabolic shift

Another interesting discovery was that there is a metabolic shift when *S. tenebrarius* is grown on glucose and the cells pass from the growth phase to the production phase. A change in metabolism has previously been observed in several other streptomycetes (Table 9-3). It is apparent that the cells adjust their metabolism to the change in function: from growth to secondary metabolite production.

Table 9-3. Metabolic shift in actinomycetes on transition from growth to secondary metabolism

Organism	Metabolic shift	Comments
S. lividans[24]	PP ↓ EMP ↑	
S. noursei[14]	PP ↓ TCA ↑	
Nonomuraea sp. ATCC 39727[10]	ED ↓ PP ↑ EMP ↑	Observed only during phosphate limitation
S. tenebrarius	ED ↓ PP ↑ EMP ↑	Growth on glucose

Some common observations can be made for metabolic shifts in *Streptomyces* that use the EMP pathway and the PP pathway. For *Streptomyces lividans*, which produces the antibiotics actinorhodin and undecylprodigiosin, it has been observed that for decreasing specific growth rates the activity of the PP pathway decreases and visa versa[24]. A similar observation of a decreased PP flux during at low specific growth rates and production of nystatin in *Streptomyces noursei* has also been reported[14].

For two antibiotic-producing actinomycetes that use the ED pathway (*Nonomuraea* spp. ATCC 39727 and *S. tenebrarius*) a common trait is that the activity of this pathway decreases after transition from growth to production, and so a larger part of the glucose (maybe even all) is metabolized via the EMP and PP pathways.
The PP and ED pathways both produce reduction equivalents in the form of NADPH. Because NADPH is necessary for growth, it is logical that the supplying pathways are mostly active during the growth period. When cell growth slows down, the need for this cofactor decreases and the energetically more favorable EMP pathway can be used.

9.5.3 Cofactor requirement for antibiotics synthesis

Could cofactor supply be the limiting factor for antibiotics production? NADPH is regarded as the cofactor necessary for biosynthesis of several antibiotics: β-lactams, polyketides and glycopeptides. NADPH is produced in the PP and ED pathways, but not in the EMP pathway. In overproducing strains of *S. lividans* the flux through the PP pathway during production phase was higher than in strains with lower antibiotics production[24]. Increased activity of the PP pathway has also been observed in connection with higher productivity of the polyketide avermectin in *S. avermitilis*[13] and the cyclopentanone antibiotic methylenomycin in *S. coelicolor* A3(2)[23].
The aminoglycosides produced by *S. tenebrarius* are only slightly more reduced than glucose, from which they are made. Tobramycin and apramycin each have one hydroxyl group reduced during biosynthesis. However, aminoglycosides need glutamate as a source of amino groups for the amination reactions. Glutamate is produced in the cell by amination of α-ketoglutarate with concurrent consumption of NADPH. If the reducing equivalents used both for uptake of ammonia (5 mol per mol of nebramycin component) and for reduction of the hydroxyl group (1 mol per mol of tobramycin or apramycin) are taken into consideration, the reducing equivalent requirement will be about 10.6 μmol per mg nebramycin (average molecular mass is 520 $g \cdot mol^{-1}$). The requirements of reducing cofactors for antibiotics synthesis will therefore be about 20 $\mu mol \cdot g^{-1}$ $DW \cdot h^{-1}$ in the cultivation with glucose as carbon source and 40 $\mu mol \cdot g^{-1}$ $DW \cdot h^{-1}$ in the cultivation with glucose/glycerol as carbon source. These requirements represent less than 2-3% of the total cofactor production flux, and it is therefore not likely that cofactor supply for antibiotics biosynthesis is limiting. On the other hand an increased ratio of $NADPH:NADP^+$ in the glucose/glycerol cultivation could in principal work as a selective pressure for production of more reduced antibiotics (tobramycin and apramycin).

9.6 Update

During preparation of this manuscript the tobramycin biosynthetic cluster has been isolated from *S. tenebrarius*[15]. Two genes were identified as possible candidates for paromamine reduction, which is the decisive step directing the aminoglycosides flux toward tobramycin instead of kanamycin. Overexpression of this reductase could shift the balance from kanamycin B towards tobramycin production.

9.7 Acknowledgements

This work was financially supported by the Danish Ministry and Alpharma ApS through a stipend to I. B. The authors are grateful to Nina Gunnarsson for scientific discussions during the course of the work. We would also like to acknowledge the

advice and help of our colleagues Søren Bendiksen, Hans Peter Smits, Michael Lynge Nielsen, Thomas Grotkjær and Mats Åkesson.

9.8 References

1. Bai, N. J., M. R. Pai, P. S. Murthy, and T. A. Venkitasubramanian. 1976. Pathways of glucose catabolism in *Mycobacterium smegmatis*. Can. J. Microbiol. 22:1374-1380.
2. Bramwell, H., H. G. Nimmo, I. S. Hunter, and J. R. Coggins. 1993. Phosphoenolpyruvate carboxylase from *Streptomyces coelicolor* A3(2): purification of the enzyme, cloning of the *ppc* gene and over-expression of the protein in a streptomycete. Biochem. J. 293:131-136.
3. Budgen, N. and M. J. Danson. 1986. Metabolism of glucose via a modified Entner-Doudoroff pathway in the thermoacidophilic archaebacterium *Thermoplasma acidophilum*. FEBS Lett. 196:207-210.
4. Christensen, B. and J. Nielsen. 1999. Isotopomer analysis using GC-MS. Metab. Eng. 1:282-290.
5. Christensen, B. and J. Nielsen. 2000. Metabolic network analysis of *Penicillium chrysogenum* using ^{13}C-labeled glucose. Biotechnol. Bioeng. 68:652-659.
6. Christensen, B. and J. Nielsen. 2000. Metabolic network analysis. A powerful tool in metabolic engineering. Adv. Biochem. Eng Biotechnol. 66:209-231.
7. Conway, T. 1992. The Entner-Doudoroff pathway: history, physiology and molecular biology. FEMS Microbiol. Rev. 9:1-27.
8. Dekleva, M. L. and W. R. Strohl. 1988. Activity of phosphoenolpyruvate carboxylase of an anthracycline-producing streptomycete. Can. J. Microbiol. 34:1241-1246.
9. Entner, N. and M. Doudoroff. 1952. Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. J. Biol. Chem. 196:853-862.
10. Gunnarsson, N., U. H. Mortensen, M. Sosio, and J. Nielsen. 2004. Identification of the Entner-Doudoroff pathway in an antibiotic-producing actinomycete species. Mol. Microbiol. 52:895-902.
11. Higgins, C. E. and R. E. Kastner. 1967. Nebramycin, a new broad-spectrum antibiotic complex. II. Description of *Streptomyces tenebrarius*. Antimicrobial. Agents Chemother. 7:324-331.
12. Hodgson, D. A. 2000. Primary metabolism and its control in *Streptomycetes*. Advances in Microbial Physiology 42:47-238.
13. Ikeda, H., H. Kotaki, H. Tanaka, and S. Omura. 1988. Involvement of glucose catabolism in avermectin production by *Streptomyces avermitilis*. Antimicrob. Agents Chemother. 32:282-284.
14. Jonsbu, E., B. Christensen, and J. Nielsen. 2001. Changes of *in vivo* fluxes through central metabolic pathways during the production of nystatin by *Streptomyces noursei* in batch culture. Appl. Microbiol. Biotechnol. 56:93-100.
15. Kharel, M. K., D. B. Basnet, H. C. Lee, K. Liou, J. S. Woo, B. G. Kim, and J. K. Sohng. 2004. Isolation and characterization of the tobramycin biosynthetic gene cluster from *Streptomyces tenebrarius*. FEMS Microbiol. Lett. 230:185-190.
16. Kieser, T., M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood. 2000. General introduction to actinomycete biology, p. 1-42. Practical *Streptomyces* Genetics. The John Innes Foundation, Norwich.
17. Koch, K. F., F. A. Davis, and J. A. Rhoades. 1973. Nebramycin: separation of the complex and identification of factors 4, 5, and 5'. J. Antibiot. (Tokyo) 26:745-751.

18. Koch, K. F. and J. A. Rhoades. 1970. Structure of nebramycin factor 6, a new aminoglycosidic antibiotic. Antimicrobial. Agents Chemother. 10:309-313.
19. Lessie, T. G. and P. V. Phibbs, Jr. 1984. Alternative pathways of carbohydrate utilization in pseudomonads. Annu. Rev. Microbiol. 38:359-388.
20. Marx, A., A. A. de Graaf, W. Wiechert, L. Eggling, and H. Sahm. 1996. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. Biotechnol. Bioeng. 49:111-129.
21. Nielsen, J. 1997. Primary metabolism, p. 63-138. Physiological engineering aspects of *Penicilium chrysogenum*. World Scientific, London, UK.
22. O'Connor, S., L. K. Lam, N. D. Jones, and M. O. Chaney. 1976. Apramycin, a unique aminocyclitol antibiotic. J. Org. Chem. 41:2087-2092.
23. Obanye, A. I., G. Hobbs, D. C. Gardner, and S. G. Oliver. 1996. Correlation between carbon flux through the pentose phosphate pathway and production of the antibiotic methylenomycin in *Streptomyces coelicolor* A3(2). Microbiology (Reading, Engl.) 142:133-137.
24. Rossa, A. C., J. White, A. Kuiper, P. W. Postma, M. Bibb, and M. J. Teixeira de Mattos. 2002. Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*. Metab Eng 4:138-150.
25. Stark, W. M., N. G. Knox, and R. M. Wilgus. 1971. Strains of *Streptomyces tenebrarius* and biosynthesis of nebramycin. Folia Microbiol. (Praha) 16:205-217.
26. Steinbuchel, A. 1986. Expression of the *Escherichia coli pfkA* gene in *Alcaligenes eutrophus* and in other gram-negative bacteria. J. Bacteriol. 166:319-327.
27. Wiechert, W. and A. A. de Graaf. 1996. In vivo stationary flux analysis by 13C labeling experiments. Adv. Biochem. Eng Biotechnol. 54:109-154.
28. Wiechert, W., M. Mollney, N. Isermann, M. Wurzel, and A. A. de Graaf. 1999. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. Biotechnol. Bioeng. 66:69-85.

APPENDICES

Appendices for chapter 3

Appendix 3.1. List of genes included in the model

D:/Appendices/Appendices for chapter 3/Appendix 3.1. List of genes included in the model.xls

Appendix 3.2. List of reactions

D:/Appendices/Appendices for chapter 3/ Appendix 3.2. List of reactions.xls

Appendix 3.3. Metabolites abbreviations

D:/Appendices/Appendices for chapter 3/ Appendix 3.3. Metabolites abbreviations.xls

Appendix 3.4. Genome comparison of S. coelicolor A3(2) and S. avermitilis

D:/Appendices/Appendices for chapter 3/ Appendix 3.4. Genome comparison of S. coelicolor A3(2) and S. avermitilis.xls

Appendix 3.5. Biomass composition of S. coelicolor A3(2)

D:/Appendices/Appendices for chapter 3/ Appendix 3.5. Biomass composition of S. coelicolor A3(2).xls

Appendix 3.6. Simulation details for the model

D:/Appendices/Appendices for chapter 3/ Appendix 3.6. Simulation details for the model.xls

Appendix 3.7. Simulation details for the non-redundant model (no isoenzymes)

D:/Appendices/Appendices for chapter 3/Appendix 3.7. Simulation details for the non-redundant model (no isoenzymes).xls

Appendix 3.8. Analysis of reactions activity

D:/Appendices/Appendices for chapter 3/ Appendix 3.8. Analysis of reactions activity.xls

Appendix 3.9. Analysis of reactions essentiality

D:/Appendices/Appendices for chapter 3/ Appendix 3.9. Analysis of reactions essentiality.xls

Appendix 3.10. List of reactions without ORFs assigned

D:/Appendices/Appendices for chapter 3/ Appendix 3.10. List of reactions without ORFs assigned.xls

Appendices for chapter 4

Appendix 4.1. S. scabies pathway and genome database in Pathway Tools format

D:/Appendices/Appendices for chapter 4/Appendix 4.1. S. scabies pathway and genome database in Pathway Tools format

Appendix 4.2. Bioopt input for S. scabies metabolic model

D:/Appendices/Appendices for chapter 4/Appendix 4.2. Bioopt input for S. scabies.txt

Appendix 4.3. Pathway holes in S. scabies Pathway Genome Database

D:/Appendices/Appendices for chapter 4/Appendix 4.3. Pathway holes in S. scabies Pathway Genome Database.xls

Appendix 4.4. Candidate genes for filling in pathway holes in S. scabies Pathway Genome Database

D:/Appendices/Appendices for chapter 4/Appendix 4.4. Candidate genes for filling in pathway holes in S. scabies Pathway Genome Database.xls

Appendix 4.5. BLAST of S. turgidiscabies pathogenicity island genes against S. scabies genes

D:/Appendices/Appendices for chapter 4/Appendix 4.5. BLAST of S. turgidiscabies pathogenicity island genes against S. scabies genes.txt

Appendices for chapter 5

Appendix 5.1. Biomass composition of S. coelicolor

D:/Appendices/Appendices for chapter 5/Appendix 5.1. Biomass composition of S. coelicolor.xls

Appendix 5.2. Loop design of microarray hybridization

D:/Appendices/Appendices for chapter 5/Appendix 5.2. Loop design of microarray.xls

Appendix 5.3. C13 labeling data

D:/Appendices/Appendices for chapter 5/Appendix 5.3. C13 labeling data.xls

Appendix 5.4. Estimation of fluxes

D:/Appendices/Appendices for chapter 5/Appendix 5.4. Estimation of fluxes.xls

Appendix 5.5. Gene expression data

D:/Appendices/Appendices for chapter 5/Appendix 5.5. Gene expression data.xls

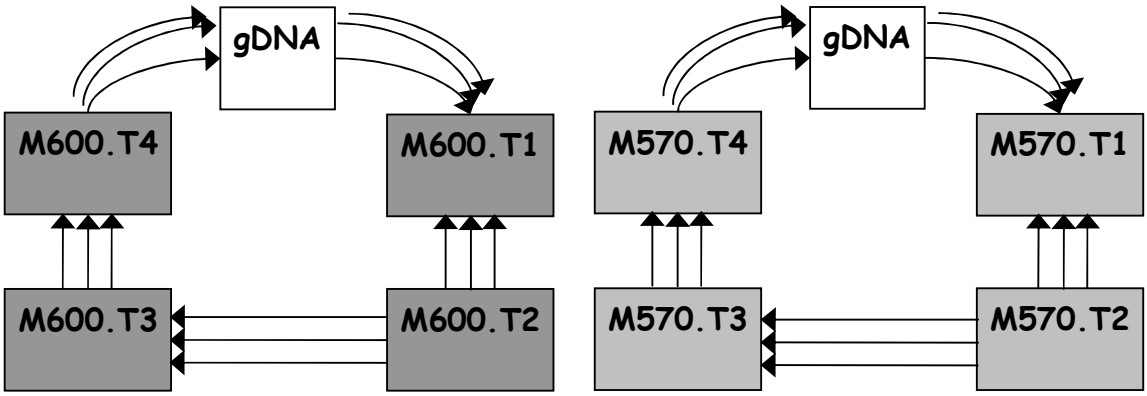
Appendix 5.6. Significantly regulated metabolic subnetwork

D:/Appendices/Appendices for chapter 5/Appendix 5.6. Significantly regulated metabolic subnetwork.xls

Appendices for chapter 6

Appendix 6.1. Design of sample hybridization on microarrays with genomic DNA as common reference

gDNA – genomic DNA; M600.T_i – RNA sample of M600 strain taken at time point *i*; M570.T_j - RNA sample of M570 strain taken at time point *j*. Each arrow corresponds to hybridization on one array, the arrow starts at Cy3-labeled probe and ends at Cy5-labeled probe.



Appendices for chapter 8

Appendix 8.1. Correlation between biomass dry weight and OD_{600} in S. coelicolor

The samples were taken from shake flask cultivations of *S. coelicolor* A3(2) M145 strain and its seven congenic mutants. There were three biological replicates. All the samples were taken at the mid exponential growth phase.